

UNIVERSITÉ LIBRE DES PAYS DES GRANDS LACS  
FACULTÉ DES SCIENCES ET TECHNOLOGIES  
DÉPARTEMENT DE GÉNIE ÉLECTRIQUE ET  
INFORMATIQUE



BP. 368 GOMA  
[www.ulpgl.net](http://www.ulpgl.net)

---

**SURVEILLANCE INTELLIGENTE DES  
SYSTÈMES PHOTOVOLTAÏQUES :  
APPROCHE BASÉE SUR LE  
MACHINE LEARNING POUR LA  
DÉTECTION D'ANOMALIES**

---

**Par :**

**CIKWANINE MASTAKI Anuarite**

Travail présenté en vue de l'obtention du Diplôme de licence en Sciences de l'Ingénieur  
**Mention** : Génie Electrique

**Directeur** : Prof. BARAKA MUSHAGE Olivier

**Encadreur** : M. Eng. MOLO MBASA Joaquim

Année Académique 2024–2025

# Epigraphe

---

*« Le meilleur moyen de prédire l'avenir est de l'inventer. »*

**Alan Kay**

# Dédicace

---

*À ces âmes courageuses qui, malgré le poids du silence et de la douleur, continuent de  
marcher avec résilience.*

***CIKWANINE MASTAKI Anuarite***

# Remerciements

---

Nous remercions le Tout Puissant pour le souffle de vie et l'accompagnement qu'il ne cesse de nous accorder.

Nos remerciements s'adressent ensuite à l'ULPGL qui nous a formés ces dernières années.

Nous exprimons notre reconnaissance la plus sincère à notre directeur, le Professeur BARAKA MUSHAGE Olivier ainsi qu'à notre encadreur le Doctorant MOLO MBASA Joaquim, pour le suivi et la patience manifestés durant ce travail.

Nous tenons à remercier tout particulièrement nos parents MASTAKI LUKOLOLA Marc et FURAHA BAYONGWA Josiane pour l'amour inconditionnel, le soutien indéfectible et les sacrifices consentis à notre égard.

Nos remerciements vont également aux sœurs de la Résurrection, à l'AFIAFEV, à RIO, à SYP, à mon oncle Etienne KASHANGABUYE et à ma tante Espérance MIHIGO pour leur soutien précieux.

Nous remercions tous ceux avec qui nous avons partagé les moments les plus marquants, particulièrement à Gérard ASHUZA, Marthe ANSIMA, Julie BUNGUKE, Emmanuella BULANGALIRE, Gilles AMPIRE, Martine IRAGI, Samuel NTWALI, Clément MIZINGWA, Jessica TCHIKO et Martine NZALAMINGI.

Nous remercions enfin nos amis et camarades, en particulier Josué SHALUKOMA, Rosalie MUHYA, Aimerance DUSHIME, Djofina VIVUYA, Alain CIBAZIGIRA, Avana DYCKOBA, Daniel KADURHA, Noces-Léonie AMPIRE, Jonathan SHAURI, Patrick TSONGO, Paul KASANANI, Jacques TSONGO, Christian TEGEJO, Cédric KASAYI, Anuarite MUSIMWA, Adonis RWABIRA, Gaël LYANI, Yvon AMPIRE, Romeo BISIMWA, Helene MUBINZA, ainsi que toutes les autres personnes bienveillantes que nous n'avons pas pu citer.

**CIKWANINE MASTAKI Anuarite**

# Résumé

---

Dans un contexte marqué par la transition énergétique et le développement croissant des installations photovoltaïques, la fiabilité et la continuité de production constituent des enjeux majeurs. En effet, les anomalies telles que l'ombrage partiel ou le court-circuit peuvent entraîner des pertes énergétiques importantes et accélérer la dégradation des équipements. D'où la nécessité de mettre en place un système de détection automatique et précoce des défauts capable de faciliter la maintenance et d'améliorer les performances des systèmes PV. Dans ce travail, notre objectif est de développer un système de détection d'anomalies basé sur les techniques de Machine Learning. Pour cela, nous avons modélisé et simulé sous MATLAB/Simulink un système photovoltaïque hors réseau intégrant un générateur PV, un convertisseur DC/DC commandé par un algorithme MPPT (P&O), une batterie lithium-ion, un convertisseur bidirectionnel et un système d'acquisition de données. Les simulations ont permis de générer un dataset annoté couvrant trois états de fonctionnement : normal, ombrage partiel et court-circuit. Nous développons un système intelligent de détection d'anomalies photovoltaïques en comparant les modèles SVM, Arbre de Décision et Random Forest. Après une simulation sous MATLAB/Simulink et préparation des données, le modèle ainsi entraîné démontre une exactitude atteignant 99,86 % pour le modèle Random Forest. L'intégration des grandeurs électriques et environnementales permet ainsi d'identifier avec précision l'ombrage partiel et les court-circuit.

**Mots clés :** Systèmes photovoltaïques, Détection d'anomalies, Machine Learning, Maintenance prédictive, Monitoring intelligent.

# Abstract

---

In a context marked by the energy transition and the growing deployment of photovoltaic installations, reliability and continuity of power generation have become major challenges. Indeed, anomalies such as partial shading or short circuits can lead to significant energy losses and accelerate equipment degradation. Hence, there is a need to implement an automatic and early fault detection system capable of facilitating maintenance and improving the performance of PV systems. In this work, our objective is to develop an anomaly detection system based on Machine Learning techniques. To achieve this, we modeled and simulated an off-grid photovoltaic system using MATLAB/Simulink, integrating a PV generator, a DC/DC converter controlled by an MPPT (P&O) algorithm, a lithium-ion battery, a bidirectional converter, and a data acquisition system. The simulations enabled the generation of a labeled dataset covering three operating conditions: normal operation, partial shading, and short circuit. We developed an intelligent photovoltaic anomaly detection system by comparing SVM, Decision Tree, and Random Forest models. After simulation using MATLAB/Simulink and curation of data, the train model demonstrates remarkable performance, with an accuracy reaching 99.86 % for the Random Forest model. The integration of electrical and environmental parameters allows precise identification of partial shading and short-circuit faults.

**Keywords:** Photovoltaic systems, Anomaly detection, Machine Learning, Predictive maintenance, Intelligent monitoring.

# Table des matières

---

<b>Remerciements</b>	<b>iii</b>
<b>Résumé</b>	<b>iv</b>
<b>Abstract</b>	<b>v</b>
<b>Table des figures</b>	<b>x</b>
<b>Liste des tableaux</b>	<b>xiii</b>
<b>Liste des abréviations</b>	<b>xiv</b>
<b>Introduction générale</b>	<b>2</b>
0.1 Contexte . . . . .	2
0.2 Identification et formulation du problème . . . . .	3
0.3 Questions de recherche . . . . .	4
Questions de recherche . . . . .	4
0.4 Formulation des hypothèses . . . . .	4
0.5 Justification du choix du sujet et motivations . . . . .	5
0.6 Énoncé des objectifs de recherche . . . . .	6
0.6.1 L'objectif général . . . . .	6
0.6.2 Les objectifs opérationnels . . . . .	6

---

0.7	Méthode et délimitation du travail . . . . .	6
0.8	Subdivision du travail . . . . .	7
<b>1</b>	<b>Généralités sur les systèmes PV, Machine Learning</b>	<b>8</b>
1.1	Introduction . . . . .	8
1.2	Système photovoltaïque . . . . .	8
1.2.1	Définition du système Photovoltaïque . . . . .	8
1.2.2	Eléments constitutifs d'un système PV . . . . .	9
1.2.3	Le module PV . . . . .	12
1.3	Diagnostiques des défauts dans les systèmes PV . . . . .	20
1.3.1	Terminologie . . . . .	21
1.3.2	Architecture du diagnostic . . . . .	22
1.3.3	Systèmes PV et ses défauts . . . . .	23
1.3.4	Détection et classification des défauts à l'aide du machine Learning	28
1.3.5	Choix des algorithmes ML à utiliser . . . . .	36
1.4	Conclusion partielle . . . . .	36
<b>2</b>	<b>Modélisation du système</b>	<b>37</b>
2.1	Introduction . . . . .	37
2.2	Conception du système (Architecture matérielle) . . . . .	37
2.2.1	Présentation du système PV étudié . . . . .	37
2.2.2	Dimensionnement du champ photovoltaïque . . . . .	38

---

2.2.3	Conversion DC/DC (Boost et MPPT) . . . . .	40
2.2.4	Gestion du stockage d'énergie . . . . .	41
2.2.5	2.2.5 Stratégie de contrôle . . . . .	42
2.2.6	2.2.6 Acquisition de données . . . . .	42
2.3	Détection d'anomalies basée sur l'intelligence artificielle . . . . .	43
2.3.1	Identification des anomalies . . . . .	43
2.3.2	Approche utilisée . . . . .	44
2.4	Technologies et outils utilisés . . . . .	49
2.4.1	MATLAB / Simulink . . . . .	49
2.4.2	Jupyter notebook . . . . .	51
2.5	Conclusion partielle . . . . .	52
<b>3</b>	<b>Implémentation et simulation du modèle d'un système PV pour la gé- nération des données</b> . . . . .	<b>53</b>
3.1	Introduction . . . . .	53
3.2	Présentation du modèle du système PV . . . . .	53
3.3	Implémentation du modèle du système PV . . . . .	54
3.3.1	Implémentation du champ PV . . . . .	54
3.3.2	Implémentation du convertisseur DC/DC . . . . .	55
3.3.3	Modèle de la batterie de stockage . . . . .	55
3.3.4	Convertisseur DC/DC bidirectionnel (type Buck-Boost) . . . . .	56

---

3.3.5	Implémentation du système d'acquisition des données électriques .	57
3.4	Simulation du système PV et génération des données . . . . .	57
3.4.1	Etiquetage des grandeurs mesurées . . . . .	57
3.4.2	Simulations des états du système . . . . .	58
3.5	Conclusion partielle . . . . .	67
<b>4</b>	<b>Détection des défauts</b>	<b>68</b>
4.1	Introduction . . . . .	68
4.2	Pipeline du modèle proposé . . . . .	68
4.3	Préparation des données . . . . .	69
4.4	Division du dataset . . . . .	69
4.5	Normalisation des données . . . . .	70
4.6	Réduction de dimension par PCA . . . . .	70
4.7	Entraînement du modèle . . . . .	71
4.8	Présentation des résultats . . . . .	71
4.9	Discussion sur les résultats . . . . .	75
4.10	Conclusion partielle . . . . .	76
	<b>Conclusion générale</b>	<b>77</b>
	<b>Bibliographie</b>	<b>79</b>
	<b>Annexes</b>	<b>84</b>

# Table des figures

---

1.1	structure d'une cellule photovoltaïque . . . . .	10
1.2	Modélisation électrique d'une cellule PV . . . . .	11
1.3	Caractéristique $I - V$ et $P - V$ d'une cellule photovoltaïque . . . . .	13
1.4	Constitution d'un champ photovoltaïque . . . . .	15
1.5	Générateur PV . . . . .	17
1.6	(a) ombrage des panneaux par une cheminée de la maison, (b) fonctionnement de la cellule dans le 3 <sup>me</sup> quadrant de la caractéristique $I - V$ , à cause de l'ombre . . . . .	18
1.7	Système photovoltaïque en : (a) fonctionnement normal, (b) présence d'une ombre : échauffement de la cellule, (c) protection par la mise en place d'une diode de bypass . . . . .	19
1.8	Différentes étapes d'un diagnostic des défauts . . . . .	22
1.9	Schéma global de détection et de diagnostic des défauts . . . . .	23
1.10	Quelques exemples de la localisation de défauts par la caméra thermique . . . . .	27
1.11	Matrice de confusion . . . . .	34
2.1	Schéma global du système choisi . . . . .	38
2.2	Ordinogramme de l'algorithme MPPT (P&O) . . . . .	41
2.3	Gestion de l'énergie . . . . .	42
2.4	Ordinogramme d'apprentissage . . . . .	44
2.5	Flow chart d'entraînement et de prédiction du SVM . . . . .	46
2.7	Flowchart de l'arbre de décision . . . . .	48
2.8	Interface de Simulink . . . . .	50
2.9	Environnement Jupyter notebook . . . . .	51
3.1	Modèle Simulink du système choisi . . . . .	54
3.2	(a) Caractéristiques du bloc PV, (b) Champ PV . . . . .	55

---

3.3	Convertisseur DC/DC . . . . .	55
3.4	Batterie de stockage (a) et ses caractéristiques (b) . . . . .	56
3.5	Convertisseur bidirectionnel Buck-Boost . . . . .	56
3.6	Bloc d'acquisition des données . . . . .	57
3.7	Etiquetage des grandeurs et marques des défauts . . . . .	58
3.8	Irradiations des modules PV en fonction du temps . . . . .	59
3.9	Variations des courants générés par les modules PV en fonction du temps	59
3.10	Tensions aux bornes des modules PV . . . . .	60
3.11	Variations des courbes du courant généré, de la tension aux bornes du champ PV et de la puissance fournie par le champ PV . . . . .	60
3.12	Variations des courbes du courant généré ou absorbée par la batterie et la tension aux bornes de la batterie . . . . .	61
3.13	Ombre partiel au niveau du module PV4 . . . . .	62
3.14	Variations des tensions aux bornes des modules avec PV ombrée . . . . .	62
3.15	Variations des tensions des modules en ombre partiel . . . . .	63
3.16	Variations du courant (a), de la tension (b) et de la puissance totale (c) durant l'ombre partiel . . . . .	63
3.17	Courbes des grandeurs de la batterie de stockage durant l'ombre . . . . .	64
3.18	Variations des tensions aux bornes des modules PV . . . . .	65
3.19	Variations des courants générés par les modules PV . . . . .	65
3.20	Variations des courants dans les strings du champ PV . . . . .	66
3.21	Variations de la tension (b), du courant (a) et de la puissance totale (c) du champ PV . . . . .	66
3.22	Variations des tensions de la batterie (a) et courant généré par la batterie (b) . . . . .	67
4.1	Pipeline du modèle . . . . .	68
4.2	Extrait statistique données d'entraînement . . . . .	70
4.3	Extrait statistique données test . . . . .	70

---

4.4	Matrice de confusion du SVM . . . . .	72
4.5	Rapport de classification du SVM . . . . .	73
4.6	Matrice de confusion de l'arbre de décision . . . . .	73
4.7	Rapport de classification du Decision Tree . . . . .	74
4.8	Matrice de confusion de Random Forest . . . . .	75
4.9	Rapport de classification du Random Forest . . . . .	75
4.10	Performances globales des modèles . . . . .	76
11	Capture d'écran du script Matlab pour la collecte des données . . . . .	84
12	Extrait de l'implémentation du modèle de Machine Learning sous Python . . . . .	85
13	Modèle Simulink du système choisi . . . . .	86

# Liste des tableaux

---

1.1	Principaux défauts et anomalies rencontrées dans un générateur PV . . .	25
1.2	Caractéristiques des modèles de ML . . . . .	31
1.3	Avantages et inconvénients des modèles étudiés . . . . .	32

# Liste des abréviations

---

**AC** : Alternating Current,

**AFNOR** : Association Française de Normalisation,

**CNN** : Convolutional Neural Network,

**DC** : Direct Current,

**ELM** : Extrême Learning Machine,

**Ff** : Facteur de forme,

**GPV** : Générateur Photovoltaïque,

**I** : Courant électrique,

**IGBT** : Insulated Gate Bipolar Transistor,

**IRM** : Indice de Rayonnement Moyen,

**ML** : Machine Learning,

**MLP** : Multi-Layer Perceptron,

**MPPT** : Maximum Power Point Tracking,

**P&O** : Perturb and Observe,

**PID** : Proportional Integral Derivative,

**PPM** : Parts Per Million,

**PV** : Photovoltaïque,

**PWM** : Pulse Width Modulation,

**RBF** : Radial Basis Function,

**RNN** : Recurrent Neural Network,

**ROC** : Receiver Operating Characteristic,

**SOC** : State of Charge,

**STD** : Standard Déviation,

**SVM** : Support Vector Machine

# Introduction générale

---

## 0.1 Contexte

Ces dix dernières années, le secteur du photovoltaïque s'est fortement développé ; cette expansion s'explique principalement par la diminution progressive des coûts de production des technologies solaires et par la mise en œuvre de politiques publiques incitatives. Ces facteurs combinés contribuent à améliorer la rentabilité des installations photovoltaïques, rendant leur retour sur investissement de plus en plus attractif pour les acteurs économiques et institutionnels [1].

Cette dynamique s'inscrit dans un contexte plus large de transition énergétique, où le photovoltaïque (PV) occupe une place de plus en plus stratégique parmi les sources d'énergies renouvelables, aux côtés de l'éolien, de l'hydroélectricité et de la biomasse. Il présente plusieurs atouts notables [2].

D'une part, l'énergie solaire est une ressource abondante et inépuisable à l'échelle humaine, le rayonnement solaire reçu chaque jour par la Terre dépassant largement les besoins énergétiques mondiaux. D'autre part, la production d'électricité photovoltaïque n'engendre aucune émission directe de CO<sub>2</sub>, ce qui en fait une technologie respectueuse de l'environnement. De surcroît, ses caractéristiques techniques, notamment son fonctionnement silencieux et sa capacité à s'adapter à différentes échelles de production, en font une solution flexible et polyvalente. Enfin, son déploiement est particulièrement avantageux pour l'électrification des zones isolées, où l'absence de réseau électrique rend les systèmes autonomes indispensables [3].

Comme tous les autres processus industriels, un système photovoltaïque peut être soumis, au cours de son fonctionnement, à différents défauts ou anomalies conduisant à une baisse de performance, voire à l'indisponibilité totale de l'installation. Toutes ces conséquences défavorables réduisent naturellement la productivité de l'installation, et donc réduit la

rentabilité de l'installation, sans compter le coût de maintenance pour remettre le système en état normal [1].

## 0.2 Identification et formulation du problème

Comme évoqué dans la section précédente, l'énergie PV s'impose aujourd'hui comme l'une des sources renouvelables incontournables. Toutefois, sur le terrain, les installations PV peuvent connaître des baisses de performance soudaines. Ce phénomène constitue un obstacle majeur à la fiabilité et de la production d'énergie, notamment dans les zones isolées, où les interventions humaines sont limitées. Cependant, la détection précoce des anomalies et des défauts reste un défi majeur, en raison de la complexité des données issues des systèmes PV, qui combinent des mesures électriques et environnementales souvent bruitées et variables.

De plus, les approches traditionnelles de contrôle reposent souvent sur des seuils fixes ou des diagnostics manuels, qui peuvent s'avérer insuffisants pour détecter rapidement et de manière intelligente les indices révélateurs de dysfonctionnements. De ce fait, la détection en temps réel des défaillances ou anomalies dans les systèmes PV demeure un défi technique, en particulier lorsqu'il s'agit de garantir une production stable dans des environnements à faible supervision.

Dans ce contexte, il devient nécessaire de mettre en place des technologies autonomes capables d'analyser de manière intelligente et continue les données issues des installations PV. Ces systèmes doivent pouvoir détecter automatiquement les signaux préliminaires indiquant l'apparition potentielle de dysfonctionnements ou d'anomalies, même dans des environnements complexes et variables. L'objectif est de garantir une surveillance proactive et fiable, minimisant ainsi les risques de pannes majeures et optimisant la performance globale des installations.

### 0.3 Questions de recherche

Tenant compte des problématiques évoquées, notre recherche s’articule autour de la question principale suivante : Quelle approche technologique permettrait d’assurer une détection intelligente des anomalies dans les installations PV à partir de données de fonctionnement d’une centrale PV ?

Pour approfondir cette interrogation, nous avons défini quelques questions spécifiques :

1. Quels indicateurs mesurables constituent des signaux préliminaires fiables de dysfonctionnement dans les systèmes photovoltaïques ?
2. Dans quelle mesure les méthodes de Machine Learning pourraient-elles améliorer la capacité d’un système à identifier automatiquement les signaux préliminaires de dysfonctionnements à partir des données électriques et environnementales d’un système PV ?
3. Quels critères ou types de données (features) contribueraient le plus à améliorer la performance d’un système de détection d’anomalies basé sur le Machine Learning pour la surveillance des installations PV ?

### 0.4 Formulation des hypothèses

À la suite des questions posées, nous formulons les hypothèses suivantes :

- La prise en compte simultanée des données électriques (tension, courant, puissance) et des variables environnementales (irradiation solaire, température, humidité) permettrait d’améliorer significativement la capacité du système à détecter précocement les anomalies.
- L’extraction et l’utilisation d’indicateurs physiques issus des données brutes permettraient d’augmenter la robustesse des modèles face aux variations saisonnières et aux bruits.

- L'intégration de caractéristiques dérivées telles que les tendances temporelles permettraient d'augmenter la robustesse et la précision des modèles de Machine Learning dans la détection des défauts photovoltaïques.

## 0.5 Justification du choix du sujet et motivations

Pour garantir un bon fonctionnement d'un système PV, il faut au minimum assurer sa maintenance. Et pourtant, la maintenance représente souvent un coût élevé, surtout lorsqu'elle intervient tardivement, une fois que les défauts ont déjà causé des pertes de production significatives.

Si certaines interventions sont planifiées, beaucoup restent imprévues et doivent être réalisées en urgence suite à une panne. Ces interventions d'urgence sont non seulement coûteuses, mais elles nécessitent également des déplacements techniques, des diagnostics complexes, et parfois le remplacement de composants défectueux.

Dans ce contexte, la capacité à anticiper les défauts à partir des données disponibles apparaît comme un levier majeur pour réduire ces coûts et améliorer la rentabilité des installations photovoltaïques. C'est pourquoi nous avons choisi de concentrer notre recherche sur l'exploration de méthodes basées sur le Machine Learning, afin de détecter ou prédire de manière précoce les dysfonctionnements dans les systèmes PV.

Au-delà des aspects techniques, ce travail présente également un intérêt économique et social important. En effet, il serait utile :

- Pour les exploitants de systèmes PV, il permettrait de limiter les interruptions de service et de mieux gérer les interventions techniques ;
- Pour les décideurs, il constituerait une base pour évaluer la viabilité des projets solaires à long terme, en tenant compte des coûts liés à l'entretien.

## 0.6 Énoncé des objectifs de recherche

### 0.6.1 L'objectif général

Ce travail vise à développer et évaluer un modèle de prédiction des défauts pour systèmes PV à l'aide de Machine Learning.

### 0.6.2 Les objectifs opérationnels

Pour réaliser notre projet dans son intégralité, nous comptons :

- Identifier et analyser les défauts typiques dans les systèmes PV ainsi que les approches utilisées pour les détecter. Dans cette étude, nous nous focalisons plus sur le défaut de court-circuit pouvant survenir à certains endroits du système PV, ainsi que le défaut d'ombrage partiel car ces derniers sont les plus fréquents et facilement simulables ;
- Modéliser et simuler un cas typique de système PV afin de générer un jeu de données synthétiques représentatif des conditions normales et défectueuses, utilisable pour l'entraînement du modèle de Machine Learning ;
- Entraîner et évaluer la performance du modèle développé.

## 0.7 Méthode et délimitation du travail

Dans ce travail, nous exploiterons une méthodologie comprenant les méthodes d'analyse moyennant les techniques :

- Documentaire, elle sera utilisée pour étudier des théories existantes sur la prédiction des défauts dans les PV et le machine Learning.
- Expérimentale, c'est celle qui sera utilisée pour effectuer les différents tests en évaluant notre modèle.

Notre travail se concentrera exclusivement sur la prédiction des défauts dans les systèmes photovoltaïques, en excluant les domaines connexes tels que la sécurité ou l'analyse statistique avancée. L'étude sera limitée à la simulation d'un cas type, afin de démontrer la faisabilité et l'efficacité de la méthode choisie, sans entrer dans des développements mathématiques lourds, tout en conservant les éléments nécessaires à la compréhension technique.

## 0.8 Subdivision du travail

Excepté l'introduction et la conclusion générale, ce travail se présente comme suit :

- Au premier chapitre, Généralités sur les défauts des systèmes PV et le Machine Learning, nous passons en revue toute la théorie nécessaire à la compréhension de notre travail.
- Au deuxième chapitre, modélisation du système PV, nous décrivons la méthodologie et les outils utilisés pour créer le jeu de données annotées qui servira de base à l'entraînement et à la validation de notre algorithme de détection de défauts basé sur les SVM.
- Au troisième chapitre, simulation du modèle d'un système PV pour la génération de données, nous simulons le modèle d'un système PV pour la génération des données.
- Au quatrième chapitre, détection des défauts, nous présentons et discutons sur les résultats des modèles de ML.

# Chapitre 1

## Généralités sur les systèmes PV, Machine Learning

---

### 1.1 Introduction

Ce chapitre parle brièvement des systèmes PV, du Machine Learning, ainsi que des différentes étapes qui seront nécessaires pour atteindre l'objectif principal de ce travail. Nous y présenterons les méthodes utilisées pour la détection de défauts dans les systèmes PV, en mettant essentiellement un accent sur le Machine Learning. Pour commencer, nous allons parcourir les différents concepts clés relatifs au système PV. Ensuite, nous allons présenter les méthodes les plus remarquables lorsqu'on veut prédire les défauts. Après cela, nous parcourons les divers modèles qui nous faciliteront d'aborder aisément le modèle le plus adapté à la prédiction /détection des défauts, qui est l'objectif même de ce travail.

### 1.2 Système photovoltaïque

#### 1.2.1 Définition du système Photovoltaïque

L'énergie photovoltaïque résulte de la conversion directe du rayonnement lumineux en électricité. Le fonctionnement des systèmes photovoltaïques repose exclusivement sur la conversion directe de la lumière solaire en électricité via des semi-conducteurs, sans processus chimique de combustion ni recours à des matières fossiles. De manière plus spécifique, L'effet photovoltaïque constitue la conversion directe de l'énergie du rayonne-

ment solaire en énergie électrique au moyen de cellules généralement à base de silicium (cellule photovoltaïque). Cela distingue nettement les systèmes photovoltaïques des centrales électriques classiques fonctionnant au charbon, au pétrole ou au gaz naturel, qui génèrent des émissions importantes de  $CO_2$  et autres polluants atmosphériques[3].

## 1.2.2 Eléments constitutifs d'un système PV

### 1.2.2.1 Cellule photovoltaïque

La cellule PV ou encore photopile est le plus petit constituant d'une installation photovoltaïque. Elle est composée de matériaux semi-conducteurs et transforme directement l'énergie lumineuse en énergie électrique. Les cellules photovoltaïques sont constituées [4] :

- D'une fine couche semi-conductrice (matériau possédant une bande interdite, qui joue le rôle de barrière d'énergie que les électrons ne peuvent franchir sans une excitation extérieure, et dont il est possible de faire varier les propriétés électroniques) tel que le silicium, qui est un matériau présentant une conductivité électrique relativement bonne ;
- D'une couche anti-reflet permettant une pénétration maximale des rayons solaires ;
- D'une grille conductrice sur le dessus ou cathode et d'un métal conducteur sur le dessous ou anode.

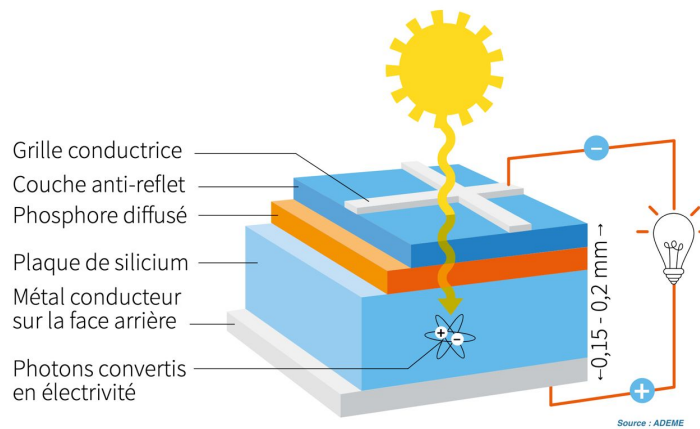


FIGURE 1.1 – Structure d’une cellule photovoltaïque [5]

Une cellule photovoltaïque est basée sur le phénomène physique appelé effet photovoltaïque qui consiste à établir une force électromotrice lorsque la surface de cette cellule est exposée à la lumière. La tension générée peut varier entre 0.3 V et 0.7 V en fonction du matériau utilisé et de sa disposition ainsi que de la température et du vieillissement de la cellule [6].

### 1.2.2.2 Modélisation d’une Cellule Photovoltaïque

Le circuit équivalent le plus simple d’une cellule solaire est une source de courant connectée en parallèle avec une diode comme le montre la Figure 1.2. Le courant délivré par la source est directement proportionnel à la lumière tombant sur la cellule. Pendant l’obscurité, la cellule solaire n’est pas active et fonctionne comme une diode. Elle ne produit ni courant ni tension. Cependant, si la lumière tombe sur la cellule solaire, elle génère un courant de diode. La diode «  $D_o$  » détermine les caractéristiques courant-tension de la cellule. Une résistance en série,  $R_s$ , représente la résistance à l’intérieur de chaque cellule, tandis que la résistance de shunt  $R_p$  est négligée car elle possède une grande valeur de résistance [7].

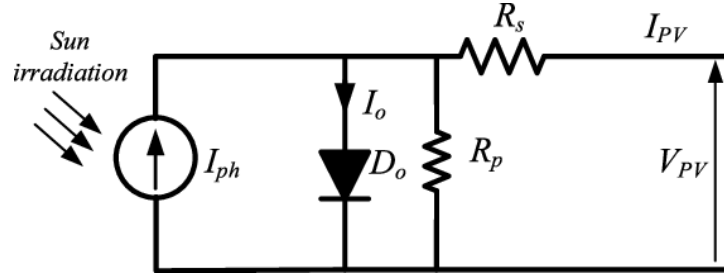


FIGURE 1.2 – Modélisation électrique d'une cellule PV [8]

Dans une cellule solaire idéale, on suppose que  $R_s = 0$  et  $R_p = \infty$ . Le courant net de la cellule est la différence entre le photo-courant  $I_{ph}$  et le courant de diode normal, qui est donné par la relation [9] :

$$I_{PV} = I_{ph} - I_0 \left( e^{\left( \frac{q(V + I \cdot R_s)}{n \cdot k \cdot T} \right)} \right) - \frac{V + I \cdot R_s}{R_p} \quad (1.0, [9])$$

Où :

- $q$  : la charge de l'électron ( $1.6 \times 10^{-19}$  C) ;
- $k$  : la constante de Boltzmann ( $1.38 \times 10^{-23}$  J/K) ;
- $T$  : la température absolue de la cellule ;
- $n$  : le facteur d'idéalité de la diode ;
- $v$  : la tension aux bornes de la cellule.

Le photo-courant  $I_{ph}$  représente le courant généré par la cellule sous l'effet du rayonnement solaire. Il dépend principalement de l'irradiance solaire incidente et, dans une moindre mesure, de la température de la cellule. À la température de référence  $T_1$ , le photo courant est proportionnel à l'irradiance solaire et peut être exprimé par :

$$I_{ph} = [I_{SC_{T_1, nom}}] \frac{G}{G_{nom}} \quad (1.1)$$

Avec :

- $G$  : le rayonnement solaire actuel ;

- $G_{nom}$  : le rayonnement solaire à la référence tester ;
- $I_{SC_{T_1, nom}}$  : le courant de court-circuit nominal de la cellule photovoltaïque à la température nominale de référence.

Lorsque la température de fonctionnement  $T$  diffère de la température de référence  $T_1$ , le photo-courant est corrigé à l'aide d'un coefficient thermique  $K_0$  :

$$I_{ph} = I_{ph} \cdot T_1 + K_0 \cdot (T - T_1) \quad (1.1, [10])$$

Le coefficient thermique  $K_0$  est déterminé à partir des courants de court-circuit mesurés à deux températures de référence  $T_1$  et  $T_2$  :

$$K_0 = \frac{I_{SC_{T_2}} - I_{SC_{T_1}}}{T_2 - T_1} \quad (1.2)$$

Le courant de saturation inverse de la diode  $I_0$  est un paramètre interne du modèle photovoltaïque qui dépend fortement de la température. Sa valeur à la température de référence  $I_{0_{T_1}}$  ne se mesure pas directement, mais elle peut être déterminée à partir des paramètres fournis par le constructeur, notamment le courant de court-circuit et la tension à vide mesurés sous conditions d'essai standardisées.  $I_0$  est donné par :

$$I_0 = I_{0_{T_1}} \left( \frac{T}{T_1} \right)^{\frac{3}{n}} e^{\left( \frac{q \cdot V_g \cdot T_1}{n \cdot k \cdot \left( \frac{1}{T} - \frac{1}{T_1} \right)} \right)} \quad (1.3)$$

Avec  $V_g$  l'énergie de bande interdite du semi-conducteur constituant la cellule en eV.

### 1.2.3 Le module PV

Un module PV est un assemblage de cellules solaires interconnectées, conçues pour convertir la lumière du soleil en électricité. Son rendement et ses performances dépendent de plusieurs facteurs techniques et environnementaux. Ces facteurs incluent :

- **Caractéristique courant-tension ( $I - V$ )** : La courbe  $I - V$  de la figure 1.3 décrit la relation entre le courant produit et la tension du module. Elle permet de déterminer la puissance maximale ( $P_{max}$ ), le point de fonctionnement optimal ( $MPP$ ), ainsi que la tension en circuit ouvert ( $V_{oc}$ ) et le courant de court-circuit ( $I_{cc}$ ).

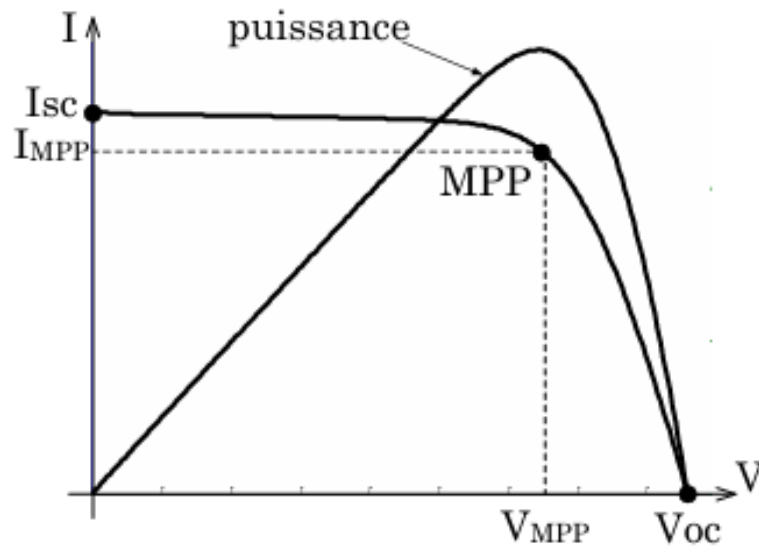


FIGURE 1.3 – Caractéristique  $I - V$  et  $P - V$  d'une cellule photovoltaïque [11]

- **Influence de l'irradiation solaire** : L'augmentation de l'ensoleillement augmente proportionnellement le courant généré. Cependant, la tension varie peu. Une bonne orientation et inclinaison des modules est donc essentielle pour maximiser la production.
- **L'influence de la température** : Une élévation de la température réduit la tension des cellules, donc diminue la puissance de sortie. C'est pourquoi le coefficient de température est un paramètre important à considérer.
- **Raccordement des modules** : Les modules sont groupés en série ou en parallèle : le groupement en série vise à augmenter la tension (somme des  $V_{oc}$ ) en gardant le courant limite au plus faible courant de la chaîne. Le groupement en parallèle augmente le courant total, mais la tension reste celle d'un seul module.

- **Rendement énergétique  $\eta$  [12]** : Le rendement d'un module est le rapport entre la puissance électrique produite ( $P_{max} = V_{max} \times I_{max}$ ) et la puissance de la radiation incidente ( $P_{inc}$ ). Il dépend de la technologie utilisée et des conditions réelles (pertes par ombrage, salissures, câblage... ) :

$$\eta = \frac{P_{max}}{P_{inc}} \quad (1.4)$$

- **Facteur de forme ( $Ff$ ) [12]** : Le facteur de forme caractérise la qualité d'un module PV. Il est défini comme le rapport entre la puissance maximale et le produit  $V_{oc} \times I_{cc}$ . Plus le  $Ff$  est élevé, meilleure est la qualité du module :

$$Ff = \frac{P_{max}}{V_{oc} \times I_{cc}} \quad (1.5)$$

### 1.2.3.1 Le string (chaîne)

La chaîne photovoltaïque est composée d'un ensemble de modules photovoltaïques branchés en série afin d'augmenter la tension de sortie du système [13].

### 1.2.3.2 Le PV Array (Champ PV)

Un *PV Array*, comme illustré à la figure 1.4, est une unité complète de production d'énergie continue. Cette unité se compose de plusieurs chaînes photovoltaïques (strings), chacune constituée du même nombre de modules. Les chaînes sont assemblées en parallèle pour augmenter le courant total, et ainsi obtenir la puissance désirée [14].

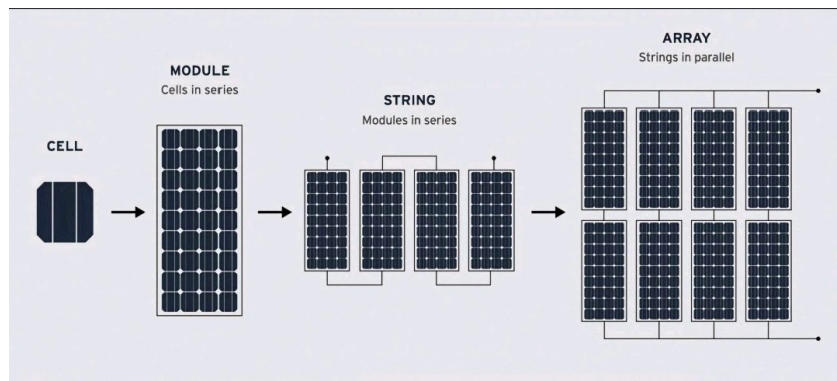


FIGURE 1.4 – Constitution d'un champ photovoltaïque [15]

### 1.2.3.3 Convertisseur Boost DC/DC

Communément appelé régulateur de charge, le convertisseur Boost DC/DC est un hacheur élévateur de tension qui assure l'interface entre le générateur photovoltaïque et la batterie. Son rôle principal est d'adapter l'impédance de la charge à celle du panneau solaire afin de forcer ce dernier à fonctionner à son point de puissance maximale (PPM), quelles que soient les variations de l'irradiation ou de la température.

Le convertisseur Boost DC/DC utilise des algorithmes qui lui permet de stabiliser le PPM; dans notre travail nous allons utiliser l'algorithme Perturbation et Observation. L'algorithme Perturbation et Observation (PO) est une méthode courante pour extraire le point de puissance maximale (MPPT) d'un panneau solaire. L'algorithme PO mesure la tension et le courant du panneau solaire, puis augmente ou diminue la tension pour observer la variation de la puissance produite. Si la puissance augmente, l'algorithme continue d'augmenter la tension jusqu'à ce que la puissance diminue. À ce stade, l'algorithme revient à la tension précédente et commence à diminuer la tension. Si la puissance diminue, l'algorithme continue de diminuer la tension jusqu'à ce que la puissance augmente. À ce stade, l'algorithme revient à la tension précédente et commence à augmenter la tension. L'algorithme PO continue de perturber la tension du panneau solaire de cette manière jusqu'à ce que la puissance maximale soit atteinte. L'algorithme PO est simple et peu coûteux, mais il peut être sensible aux variations de l'ensoleillement

et peut entraîner des pertes de puissance. Des variantes de cette méthode ont été proposées pour améliorer ses performances, telles que l'algorithme PO amélioré avec espace de recherche confiné. L'algorithme PO est l'une des méthodes les plus courantes pour extraire le point de puissance maximale d'un panneau solaire, mais il existe d'autres méthodes plus avancées qui peuvent être plus précises et plus rapides [16].

#### **1.2.3.4 Batterie de stockage**

C'est un système électrochimique réversible : il stocke de l'énergie sous forme chimique et la restitue sous forme électrique. On distingue plusieurs types de batteries notamment les batteries au nickel, les batteries au lithium et les batteries au plomb. Dans les systèmes à énergie renouvelable la batterie au plomb est la plus utilisée [17].

#### **1.2.3.5 Convertisseur Bidirectionnel Buck-Boost**

Le convertisseur DC/DC bidirectionnel Buck-Boost agit comme une interface entre la source photovoltaïque et la batterie, capable d'inverser le sens du transfert de puissance selon les besoins du système. Dans un seul et même circuit, il combine deux modes de fonctionnement : le mode abaisseur (Buck), qui réduit la tension du panneau pour charger la batterie en période de forte production, et le mode élévateur (Boost), qui augmente la tension de la batterie pour alimenter la charge lorsque le soleil fait défaut. Cette réversibilité est pilotée par une commande électronique (souvent de type PWM) qui alterne l'activation de deux interrupteurs MOSFET, permettant ainsi de maintenir la stabilité de la tension du bus continu tout en gérant intelligemment le stockage d'énergie [18].

### 1.2.3.6 Fonctionnement du générateur PV à sa puissance maximale

- Le générateur photovoltaïque La cellule photovoltaïque élémentaire constitue un générateur de très faible puissance vis-à-vis des besoins de la plupart des applications domestiques ou industrielles. Une cellule élémentaire de quelques dizaines de centimètres carrés délivre, au maximum, quelques watts sous une tension inférieure au volt typiquement liée à une tension de jonction PN. Pour produire davantage de puissance, plusieurs cellules doivent être assemblées afin de créer un module photovoltaïque complet ou générateur d'énergie [19].

Un générateur photovoltaïque est un dispositif qui convertit l'énergie solaire en électricité utilisable dont la caractéristique  $I = f(U)$  est fortement non linéaire. En conséquence, pour un même éclairement la puissance délivrée sera différente selon la résistance instantanée du circuit.



FIGURE 1.5 – Générateur PV [20]

### 1.2.3.7 Protections classiques d'un générateur PV

Les protections classiques d'un générateur photovoltaïque incluent les disjoncteurs, parafoudres, dispositifs de mise à la terre et interrupteurs sectionneurs, tandis qu'un paratonnerre peut protéger la structure contre la foudre. Les onduleurs modernes intègrent

également des protections contre les surcharges, court-circuit, variations de tension et surchauffe. Globalement, une installation PV nécessite des protections pour les intervenants, contre la foudre et pour le générateur lui-même.

Le générateur PV comme toute autre générateur d'énergie doit être protégé contre les surintensités, et autre types de problèmes dont :

**Ombre d'un générateur** L'ombrage partiel d'une cellule va forcer cette dernière à travailler dans le quadrant Q3 (voir figure 1.5(b)), c'est-à-dire d'inverser la polarité de la tension de l'élément et de l'élever au seuil tension inverse de la jonction ( $U_C \approx -15$  V à  $-25$  V). La puissance absorbée par les cellules à l'ombre dépasse très nettement la puissance normalement dissipée et provoque des points chauds.

Les points chauds peuvent endommager définitivement le module PV. Une protection contre les surintensités est sans effet, car l'augmentation de la puissance à dissiper est liée à l'apparition d'une tension inverse dans la cellule affectée et non à une augmentation significative du courant  $I_{SC}$  [21].

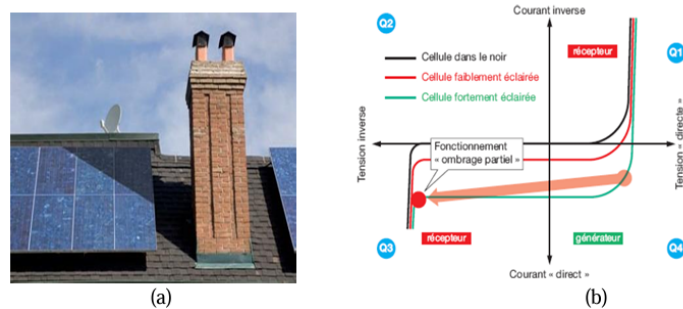


FIGURE 1.6 – (a) ombrage des panneaux par une cheminée de la maison, (b) fonctionnement de la cellule dans le 3<sup>me</sup> quadrant de la caractéristique  $I - V$ , à cause de l'ombre [22]

Une diode By-pass va, en permettant au courant des autres éléments en série de contourner la cellule « ombrée » (figure ci-dessous), permettre alors de :

- Eviter la surtension inverse ainsi que les points chauds liés à cet ombrage,

- Laisser les autres cellules non ombragées de la chaîne générer leur courant normal, à la place du courant sensiblement égal au courant réduit fourni par la cellule ombragée.

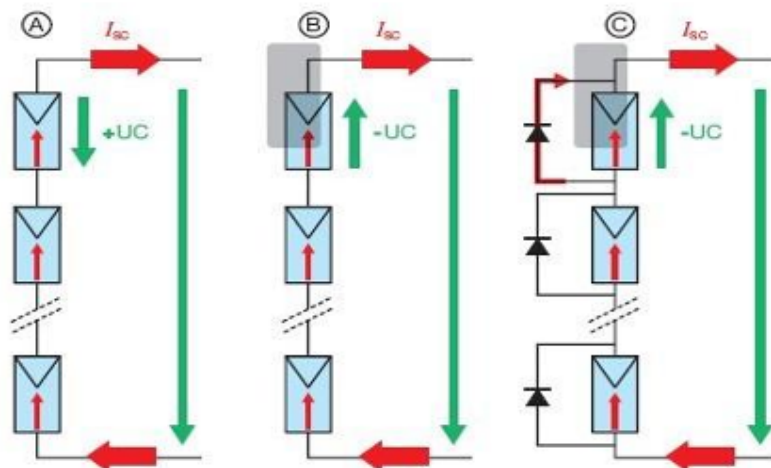


FIGURE 1.7 – Système photovoltaïque en : (a) fonctionnement normal, (b) présence d’une ombre : échauffement de la cellule, (c) protection par la mise en place d’une diode de bypass [23]

### 1.2.3.8 Application des systèmes PV [24]

Les applications de l’énergie produite par les systèmes PV varient selon le type d’installation : autonome, hybride ou connecté au réseau. Ils répondent aux besoins énergétiques en zones rurales, urbaines et industrielles.

**Installations photovoltaïques autonomes** Ces systèmes photovoltaïques sont installés pour assurer un fonctionnement autonome sans recours à d’autres sources d’énergie. Généralement, ces systèmes sont utilisés dans les régions isolées et éloignées du réseau.

**Installations photovoltaïques hybrides** Les systèmes d’énergie hybride associent au moins deux sources d’énergie renouvelable aussi qu’une ou plusieurs sources d’énergie classiques. Les sources d’énergie renouvelable, comme le photovoltaïque et l’éolienne ne

délivrent pas une puissance constante, mais vu leurs complémentarités, leur association permet d'obtenir une production électrique continue.

**Installations photovoltaïques raccordées au réseau** Un tel système s'installe sur un site raccordé au réseau. Généralement sur des habitations ou des entreprises qui souhaitent recourir à une forme d'énergie renouvelable et qui bénéficient d'un bon ensoleillement.

Un générateur photovoltaïque connecté au réseau n'a pas besoin de stockage d'énergie et élimine donc le maillon le plus problématique (et le plus cher). C'est en fait le réseau dans son ensemble qui sert de réservoir d'énergie.

### 1.3 Diagnostiques des défauts dans les systèmes PV

De nos jours, le diagnostic de systèmes est une discipline complexe et en évolution constante, qui intègre une combinaison de techniques et de technologies, allant des méthodes statistiques et informatiques traditionnelles aux approches plus récentes d'intelligence artificielle et d'apprentissage automatique. Grâce à ces outils et techniques, les ingénieurs et les techniciens peuvent rapidement identifier les anomalies dans le fonctionnement des systèmes et proposer des solutions pour les résoudre, garantissant ainsi la fiabilité, la sécurité et l'efficacité des systèmes.

**C'est quoi le diagnostic ?** Le diagnostic est le raisonnement menant à l'identification de la cause (l'origine) d'une défaillance, d'un problème ou d'une maladie, à partir de symptômes relevés par des observations, des contrôles ou des tests. Il ne doit pas être confondu avec la diagnose qui est la détermination d'une espèce biologique par rapport à une autre (c'est une des activités principales d'une science appelée la taxinomie), à partir de caractères distinctifs [25].

**C'est quoi le but de diagnostic ?** Le diagnostic d'un système vise principalement à détecter et localiser les défaillances des équipements, constituant une étape clé de la maintenance. Selon l'association Française de Normalisation (AFNOR), la maintenance regroupe toutes les actions nécessaires pour maintenir un système dans l'état lui permettant d'assurer sa fonction.

Dans un système photovoltaïque, le diagnostic permet d'identifier les anomalies affectant le rendement, assurant ainsi la performance, la fiabilité et la sécurité de l'installation tout en prévenant les pertes énergétiques et en prolongeant sa durée de vie.

**Que se passe-t-il lorsque le diagnostic n'est pas utilisé ?** L'absence de diagnostic peut avoir de lourdes conséquences : les problèmes existants peuvent passer inaperçus, compromettant la sécurité des personnes, de l'environnement et des biens. Elle peut également entraîner des coûts plus élevés à long terme et des décisions inexactes, basées sur des hypothèses plutôt que sur des faits.

### 1.3.1 Terminologie

Il existe une variété de termes utilisés dans la littérature. Nous utiliserons ce qui suit [26] :

- **Défaut** : Différence entre une caractéristique mesurée et sa valeur de référence. Un défaut peut conduire à une défaillance du système tout comme il peut exister sans altérer son fonctionnement général.
- **Défaillance** : est une conséquence possible d'une altération importante des caractéristiques. Elle correspond à l'incapacité du système à réaliser la fonction souhaitée. Elle peut être passagère ou permanente.
- **Panne** : est l'interruption permanente du système à réaliser la fonction souhaitée.
- **Résidus** : Indicateurs d'anomalie exprimant la différence entre les valeurs théoriques et celles mesurées. Il présente une valeur nulle lors du fonctionnement normal.

- **Monitoring** : c'est la surveillance en temps réel du système, il permet la détection précoce des défauts.
- **Supervision** : c'est la surveillance du système et la prise de décisions rapide dans le cas de l'existence d'un défaut.
- **Détection de défaut** : il s'agit de détecter rapidement les anomalies et de prendre une décision rapide quant à l'existence ou non d'un défaut dans le système.
- **Isolation des défauts** : il s'agit de déterminer le composant qui cause le problème, par exemple, quel capteur ou actionneur est devenu défectueux.
- **Identification du défaut** : il s'agit d'estimer la taille, le type ou la nature du défaut.

Voici les étapes à suivre pour faire un diagnostic donné dans ce diagramme :



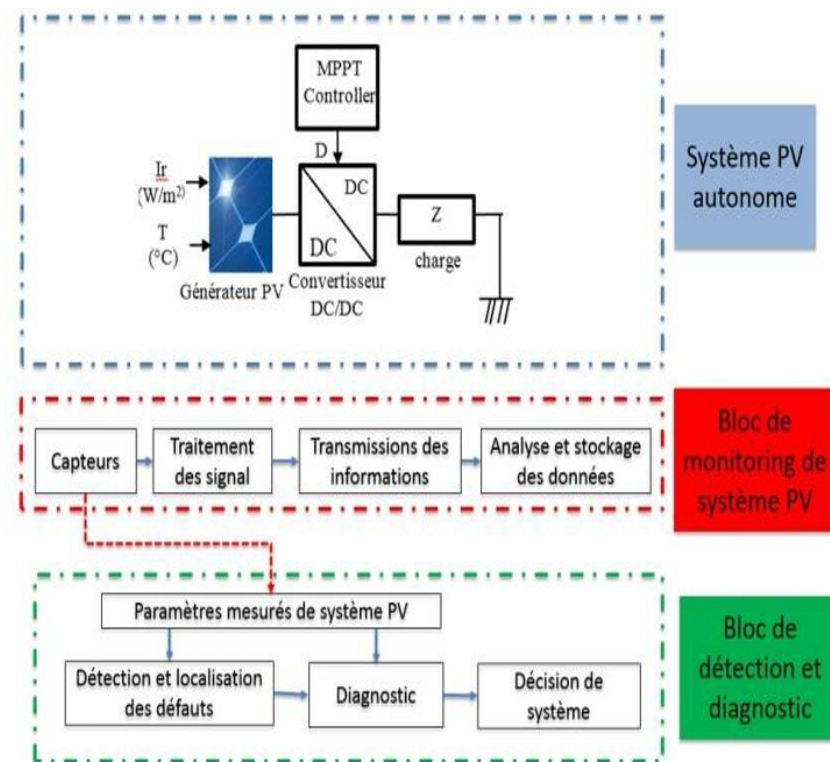
FIGURE 1.8 – Différentes étapes d'un diagnostic des défauts [27]

### 1.3.2 Architecture du diagnostic

Les principaux objectifs de la surveillance du système PV sont la détection de la défaillance, l'évaluation de la performance et l'assurance du bon fonctionnement du sys-

tème. Pour atteindre ces objectifs, des composants dédiés, tels que les capteurs, l'acquisition de données systèmes, les systèmes de communication de données, le logiciel dédié et les algorithmes pour l'analyse de données sont utilisés comme l'indique la figure 1.10.

La taille et le fonctionnement du système et des coûts d'entretien sont des paramètres qui déterminent le système de surveillance adéquat. En revanche, la taille de l'installation photovoltaïque est importante pour la conception du système de surveillance [21].



capteurs câblés dans les petites installations PV est économique et moins complexe.

FIGURE 1.9 – Schéma global de détection et de diagnostic des défauts [28]

### 1.3.3 Systèmes PV et ses défauts

Le générateur photovoltaïque est le seul convertisseur direct pour transformer la lumière en énergie électrique, et offre la possibilité de produire de l'électricité directement à partir d'une ressource renouvelable et largement disponible. Le champ d'utilisation des systèmes photovoltaïques est large avec beaucoup de configurations autonomes et reliées au réseau

de distribution. Les applications de l'énergie solaire incluent le pompage d'eau, réfrigération, climatisation, sources lumineuses, véhicules électriques, centrales photovoltaïques, utilisation militaire, domaine spatial, et dans les systèmes hybrides.

A travers ce point, nous essayons de mettre en relief l'intérêt du diagnostic à établir sur un système photovoltaïque PV en défaillance. L'objectif est d'arriver à assister le concepteur ou le réparateur à mieux cerner le problème de la défaillance affectant ce dispositif PV.

**A. Défaits dans le système photovoltaïque** Au cours de son fonctionnement, une installation PV peut être éventuellement soumise à différents défauts et conditions de fonctionnement. Les défauts et les anomalies varient d'une installation à une autre en fonction de la conception, l'installation, l'opération et la maintenance.

Une liste des origines de défauts les plus fréquents est montrée dans le tableau 1.1. Ce tableau a été établi en considérant le type du défaut, sa conséquence principale, puis son degré d'impact sur la production du système ou criticité (1 : faible, 2 : moyen, 3 : fort), son occurrence (1 : faible, 2 : moyenne, 3 : forte) ainsi que sa phase d'origine (C : Conception ; I : Installation ; E : Exploitation).

Dans ce travail, nous ne retiendrons que les défauts principaux. Le critère de sélection de ces défauts repose sur le produit entre la criticité et l'occurrence. Suite au retour d'expérience de Trans énergie (responsable du réseau de transport de l'électricité du Québec), les défauts avec un score supérieur ou égal à 3 ont été retenus et sont listés dans le Tableau 1.1 [29].

TABLE 1.1 – Principaux défauts et anomalies rencontrés dans un générateur PV

Éléments du système	Origines de défauts et d'anomalies
Générateur PV	<p>Feuilles d'arbre, déjections, pollution, sable, neige etc.</p> <p>Détérioration des cellules, fissure, échauffement des cellules</p> <p>Pénétration de l'humidité, dégradation des interconnexions, corrosion des liaisons entre les cellules</p> <p>Modules de performances différentes</p> <p>Module arraché ou cassé</p> <p>Modules court-circuités</p> <p>Modules inversés</p>
Boîte de jonction	<p>Rupture du circuit électrique</p> <p>Court-circuit du circuit électrique</p> <p>Destruction de la liaison</p> <p>Corrosion des connexions</p>
Câblage et connecteur	<p>Circuit ouvert</p> <p>Court-circuit</p> <p>Mauvais câblage (module inversé)</p> <p>Corrosion des contacts</p> <p>Rupture du circuit électrique</p>
Diode de protection (diode de bypass et diode anti-retour)	<p>Destruction des diodes</p> <p>Absence ou non fonctionnement de diodes</p> <p>Inversion de la polarité des diodes au montage, diode mal connectée</p>

### 1.3.3.1 Méthode de diagnostic d'un champ PV

Dans la partie précédente, les principaux défauts ont été retenus à partir de la liste des défauts les plus fréquents en se fondant sur leur criticité et occurrence. Dans cette partie, nous montrons les méthodes de diagnostic existantes, même si l'intégralité des défauts retenus n'y est pas prise en compte. Lors de la recherche des méthodes de diagnostic, il faut différencier ses deux fonctions : la détection et la localisation. Certaines méthodes utilisées ou proposées effectuent uniquement la fonction de détection. Tandis que certaines autres effectuent la fonction de localisation après que la détection de défauts a été réalisée. Quelques fonctionnalités sont aussi prises en compte telles que la finesse de diagnostic, les mesures nécessaires et la capacité de diagnostiquer en ligne.

**Méthodes non-électriques** Il existe plusieurs méthodes non-électriques, destructives ou non destructives, pour diagnostiquer le défaut au niveau de cellule PV. Le défaut principal qui peut avoir lieu à ce niveau est la fissure de la cellule. On peut citer comme méthodes : les essais mécaniques de flexion, l'imagerie par photoluminescence et électroluminescence, tests de thermographie. Au niveau du module PV, la méthode de l'imagerie d'infrarouge (caméra thermique) est largement appliquée, exemple fig. 1.10. Cette méthode repose sur le fait que tous les matériels émettent un rayonnement infrarouge sur une plage de longueur d'onde qui dépend de la température du matériau. En examinant la distribution de la température au niveau du module, des anomalies (si elles ont lieu) peuvent être localisées [30].

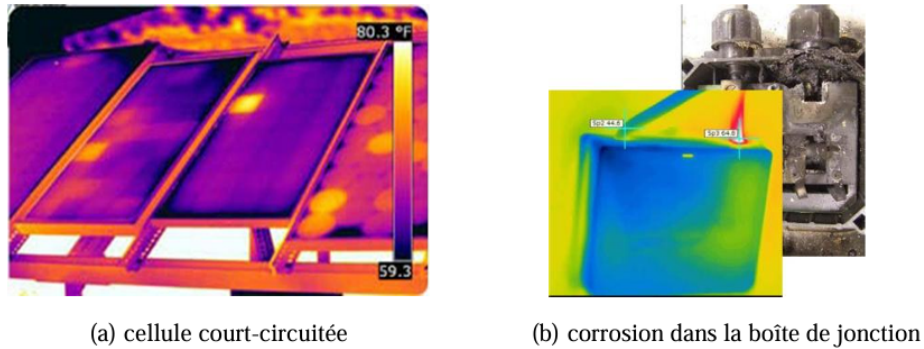


FIGURE 1.10 – Quelques exemples de la localisation de défauts par la caméra thermique [31]

**Méthodes électriques** Les systèmes de monitoring actuels d'un système PV sont essentiellement intégrés aux onduleurs. Dans ce cadre, les données mesurées sont souvent les mêmes d'un système à l'autre. Les grandeurs mesurées les plus courantes sont :

- Le courant débité par le champ PV
- La tension aux bornes du champ PV
- La résistance d'isolement entre les bornes positive et négative du champ PV

Il est aussi possible d'ajouter les grandeurs complémentaires que sont la température ambiante du site et l'ensoleillement aux mesures électriques. Ces données nécessitent un capteur spécifique (sonde de température et cellule de référence). Il convient, dans le cas où ces grandeurs sont nécessaires, de les traiter à l'aide d'un automate spécifique. Cet automate est relié à l'onduleur et centralise la totalité des données afin de les enregistrer et/ou de les envoyer sur un serveur distant. Les mesures du côté AC sont plus importantes en nombre car directement liées à l'énergie qui sera vendue. Il est courant de relever :

- Le courant AC ;
- La tension AC ;
- la fréquence ;
- L'impédance du réseau vue par l'onduleur

Des mesures décrites dans les deux paragraphes précédents, il est aisé de déduire : La

puissance instantanée DC, La puissance instantanée AC, L'énergie produite sur différentes périodes (suivant la capacité de stockage des données) côtés DC et AC.

### 1.3.4 Détection et classification des défauts à l'aide du machine Learning

Les algorithmes d'apprentissage automatique sont regroupés en quatre catégories dans la littérature, à savoir :

- l'apprentissage supervisé, pour chaque observation, il existe une étiquette à laquelle le modèle d'apprentissage automatique ajusté apprend à faire correspondre les prédicteurs dans le but soit de prédire les étiquettes pour des observations futures non vues, soit de comprendre la relation entre les prédicteurs et l'étiquette [32].
- L'apprentissage non supervisé, ici il n'y a que des variables prédictives pour chaque observation. L'absence d'étiquettes associées crée un environnement relativement aveugle, et les modèles d'apprentissage automatique sont utilisés pour extraire des relations soit entre les prédicteurs, soit entre les observations afin de fournir des informations utiles sur le problème [32]
- L'apprentissage semi-supervisé présente divers contextes d'application pour les problèmes d'apprentissage supervisé et non supervisé. L'une des applications les plus courantes des techniques d'apprentissage semisupervisé est la classification semi-supervisée lorsque seule une petite quantité de données étiquetées et une grande quantité de données non étiquetées sont disponibles. L'étiquetage pseudo des données non étiquetées comme première étape et leur utilisation avec les données étiquetées comme deuxième étape devrait permettre d'obtenir de meilleures performances que l'utilisation des seules données étiquetées.
- L'apprentissage par renforcement repose sur l'apprentissage issu de l'interaction, ce qui le distingue des autres approches d'apprentissage automatique. Il s'agit d'une méthode d'apprentissage orientée vers un objectif qui apprend à partir d'interactions avec un environnement donné. L'apprentissage par renforcement vise à ap-

prendre des correspondances entre les situations et les actions afin de maximiser un signal de récompense à long terme. Le processus d'apprentissage par interactions nécessite de trouver un équilibre entre exploration et exploitation, ce qui constitue un défi propre à l'apprentissage par renforcement [33].

#### 1.3.4.1 Technologies de Machine Learning

- **Régression** : La régression est une technique d'apprentissage automatique supervisé dont l'objectif est de prédire une valeur numérique continue (un nombre réel). Par exemple : prédire le prix d'une maison en fonction de sa superficie, estimer la demande future d'un produit ou prédire le rendement d'une culture agricole. Son principe est de modéliser la relation entre une ou plusieurs variables explicatives et une variable cible continue [34].
- **Classification** : La classification est une famille de techniques d'apprentissage automatique supervisé dont l'objectif est de prédire l'appartenance d'une observation à une catégorie discrète (une classe) choisie parmi un ensemble fini de possibilités (ex : "spam" ou "non spam", "malade" ou "sain"). Les algorithmes de classification apprennent à établir des frontières de décision pour séparer les différentes classes. Contrairement à la régression, elle attribue une étiquette qualitative [34].

#### 1.3.4.2 Quelques algorithmes de Machine Learning couramment utilisés

En Machine Learning, plusieurs algorithmes sont couramment utilisés pour la prédiction et la localisation des défauts dans les systèmes photovoltaïques. Les plus courants dans ce domaine sont :

- Les Réseaux de Neurones Convolutifs (CNN) ;
- Les Réseaux de Neurones Récurrents (RNN) ;
- Les Machines à Vecteurs de Support (SVM) ;

- Les Forêts d'Arbres de Décision (Random Forest) ;
- Les Réseaux de Neurones Multicouches (MLP) ;
- Les Machines d'Apprentissage Extrême (ELM).

Bien que tous ces algorithmes visent à détecter des défauts dans les systèmes photovoltaïques, ils ne traitent pas les données de la même manière :

- **Machine Learning "Classique" (SVM, Random Forest)** : Ces modèles s'appuient généralement sur l'extraction manuelle de caractéristiques (*feature engineering*). L'ingénieur doit identifier au préalable quels paramètres (tension, courant, température) sont les plus pertinents pour que l'algorithme puisse classer le défaut [35].
- **Deep Learning (CNN, RNN, MLP, ELM)** : Inspirés par la structure du cerveau humain (réseaux de neurones artificiels), ces modèles sont capables d'apprendre automatiquement les caractéristiques importantes à partir de données complexes. Ils sont particulièrement performants avec de grands volumes de données [35].

Le tableau 1.2 ci-dessous présente une synthèse comparative de leurs principes fondamentaux, de leurs types et de leurs applications typiques dans le domaine de la détection de défauts photovoltaïques [34].

TABLE 1.2 – Caractéristiques des modèles de ML

Modèle	Architecture	Caractéristique	Utilisation
CNN	Convulsive	Utilise des convolutions pour extraire des caractéristiques des données	Vision par ordinateur, traitement du langage naturel, classification d'image, segmentation d'images.
RNN	Récurrente	Prend en compte la chronologie des données	Analyse du langage naturel, traduction automatique, reconnaissance vocale.
Isolation Forest	Détection d'anomalies	Isole les points des données anormaux	Détection de défauts, détection de fraude, détection de spam.
Random Forest	Ensemble	Mixe les prédictions de plusieurs arbres de décision	Classification, régression, détection d'anomalie.
SVM	Support Vector Machine	Utilise des vecteurs de support pour séparer les classes	Classification et régression.
ELM	Extrême Machine Learning	Utilise une méthode d'apprentissage rapide	Classification, régression.
MLP	Multicouche perceptron	Utilise plusieurs couches de neurones pour apprendre des modèles complexes	Classification, régression.

Le tableau 1.3. Synthétise leurs principaux avantages et inconvénients dans le contexte de la maintenance des systèmes photovoltaïques.

TABLE 1.3 – Avantages et inconvénients des modèles étudiés

Modèle	Avantages	Inconvénients
CNN	Efficace pour la détection des motifs visuels subtils dans les données structurées telles que les images. Architecture spécifique conçue pour extraire automatiquement des caractéristiques pertinentes à partir des données d'entrée.	Exige des ensembles des données volumineux, complet et de haute qualité pour obtenir des résultats précis et fiables. Coût computationnel élevé.
RNN	Adapté au traitement des données séquentielles, telles que des séquences temporelles ou des séquences d'événements. Utiles pour analyser des données où la chronologie joue un rôle critique.	Sensibles aux problèmes de disparition du gradient. Coût computationnel élevé.
Isolation Forest	Capacité à identifier efficacement les observations aberrantes. Rapidité d'identification des points des données anormaux.	Peut être sensible aux données bruyantes. Peut-être moins efficace pour les données hautement dimensionnelles.
Random Forest	Agrégation de multiples arbres de décision pour une meilleure performance. La classification, la régression et la détection des anomalies passent.	Peuvent être sensibles aux données bruyantes. Peuvent être moins efficaces pour les données hautement dimensionnelles.

Modèle	Avantages	Inconvénients
SVM	Fait pour la classification, la régression et la détection des anomalies. Efficace pour les données linéairement ou non linéairement séparés dans un espace de grande dimension.	Peuvent être sensible aux données bruyantes. Coût computationnel élevé pour les grands ensembles de données.
ELM	Phase d'entraînement rapide. Entraînement efficace avec un cout computationnel réduit.	Peut être sensible aux données bruyantes. Peut nécessiter une sélection prudente des données pour une performance optimale.
MLP	Architecture en couches pour une meilleure performance. Adaptés à la résolution de problèmes de classification et de régression complexes.	Peuvent être sensibles aux données bruyantes. Coût computationnel élevé pour les grands ensembles de données.

### 1.3.4.3 Métriques de performance en détection de défauts par ML

L'évaluation des performances d'un modèle constitue une étape critique dans tout projet de Machine Learning, visant à déterminer sa capacité à généraliser et à prédire correctement la variable cible face à des nouvelles données non observées lors de l'entraînement.

Dans le cadre spécifique de la détection de défauts, plusieurs métriques quantitatives permettent de mesurer objectivement la qualité d'un modèle. Les indicateurs les plus couramment utilisés comprennent la sensibilité, la spécificité, la précision, le rappel, le score F1, ainsi que l'analyse de la matrice de confusion et de l'aire sous la courbe ROC (AUC-ROC) [34].

**Matrice de confusion (Confusion Matrix) [36]** La matrice de confusion est un outil de visualisation tabulaire essentiel pour évaluer un modèle de classification. Elle présente une

synthèse comparative entre les prédictions du modèle et les valeurs réelles (ground truth), en catégorisant les résultats selon quatre cas : les vrais positifs, les faux positifs, les vrais négatifs et les faux négatifs.

		Predicted Class		
		Positive	Negative	
Actual Class	Positive	True Positive (TP)	False Negative (FN) Type II Error	Sensitivity $\frac{TP}{(TP + FN)}$
	Negative	False Positive (FP) Type I Error	True Negative (TN)	Specificity $\frac{TN}{(TN + FP)}$
		Precision $\frac{TP}{(TP + FP)}$	Negative Predictive Value $\frac{TN}{(TN + FN)}$	Accuracy $\frac{TP + TN}{(TP + FP + FN + TN)}$

FIGURE 1.11 – Matrice de confusion [37]

- **Vrais Positifs (True Positives)** : Ce terme désigne les cas où le modèle détecte correctement la présence d'un défaut dans le système photovoltaïque. La prédiction positive correspond alors à l'état réel du système, par exemple lorsque le modèle signale un défaut sur un panneau et qu'une inspection confirme effectivement ce défaut.
- **Vrais Négatifs (True Negatives)** : Il s'agit des situations où le modèle identifie correctement l'absence de défaut, la prédiction négative correspondant à l'état réel normal du système, comme lorsqu'aucun défaut n'est signalé et que l'inspection le confirme.
- **Faux Positifs (False Positives)** : Les faux positifs désignent les cas où le modèle indique à tort la présence d'un défaut alors que le système est normal, entraînant ainsi une alarme injustifiée.
- **Faux Négatifs (False Negatives)** : Cette situation correspond à un cas où le modèle prédit à tort l'absence de défaut alors qu'un défaut est réellement présent. Il s'agit donc d'une prédiction négative erronée, appelée faux négatif, qui traduit un défaut non détecté par le modèle.

Bien que la matrice de confusion ne constitue pas une métrique à elle seule, elle fournit une base essentielle pour le calcul et l'interprétation des indicateurs de performance.

- **Exactitude (Accuracy) [38]** L'exactitude mesure la proportion globale de prédictions correctes (qu'elles soient positives ou négatives) par rapport au nombre total de prédictions effectuées par le modèle. Elle reflète ainsi la capacité générale du modèle à classifier correctement les instances.

$$\text{Exactitude} = \frac{\text{Nombre de prédictions correctes}}{\text{Nombres de prédictions totales}} \quad (1.6)$$

- **Sensibilité [39]** La sensibilité quantifie la capacité du modèle à détecter tous les échantillons positifs réels. Elle est donnée par la relation suivante :

$$\text{Sensibilité} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux négatifs}} \quad (1.7)$$

- **Spécificité [39]** La spécificité évalue la capacité du modèle à identifier correctement les échantillons négatifs réels. Elle se calcule par :

$$\text{Spécificité} = \frac{\text{Vrais négatifs}}{\text{Vrais négatifs} + \text{Faux positifs}} \quad (1.8)$$

- **Précision [39]** Cet indicateur mesure la fiabilité des prédictions positives du modèle, c'est-à-dire la proportion d'exemples prédits comme positifs qui le sont véritablement.

$$\text{Précision} = \frac{\text{Vrais positifs}}{\text{Vrais positifs} + \text{Faux positifs}} \quad (1.9)$$

- **Score F1 [39]** Le score F1 correspond à la moyenne harmonique de la précision et du rappel (sensibilité). Il offre ainsi une mesure unique équilibrant ces deux métriques, particulièrement utile lorsque les classes sont déséquilibrées.

$$F1score = 2 \cdot \frac{\text{Précision} \cdot \text{Rappel}}{\text{Précision} + \text{Rappel}} \quad (1.10)$$

### 1.3.5 Choix des algorithmes ML à utiliser

À l'issue de l'analyse de la littérature, il ressort que les *Support Vector Machine* (SVM) comptent parmi les algorithmes les plus fréquemment mobilisés dans le diagnostic des systèmes photovoltaïques, en raison de leur efficacité dans les problèmes de classification multi-classes et de leurs performances satisfaisantes, même avec des jeux de données de taille moyenne.

Par ailleurs, les approches fondées sur les *Decision Tree* (arbres de décision) et les *Random Forest* (forêts aléatoires) sont également largement adoptées, notamment pour leur capacité d'interprétation et leur robustesse face aux variations des données.

Ainsi, le choix des algorithmes retenus dans cette étude s'appuie principalement sur leur usage répandu et leur pertinence démontrée dans les travaux existants [38], [40], [41].

## 1.4 Conclusion partielle

Au cours de ce chapitre, nous avons présenté les éléments constitutifs de la chaîne photovoltaïque et modélisé électriquement le comportement de la cellule. Nous avons classifié les différents types de défauts et montré leurs impacts négatifs. Enfin, nous avons introduit les concepts du Machine Learning comme une solution moderne et automatisée pour le diagnostic.

Dans le chapitre suivant, nous détaillerons la méthodologie de recherche, les outils de simulation utilisés et la base de données qui servira à l'entraînement.

# Chapitre 2

## Modélisation du système

---

### 2.1 Introduction

Dans ce chapitre, nous présentons la modélisation du système que nous allons étudier. Nous commençons par la présentation du système photovoltaïque étudié ainsi que ses différents composants. Ensuite, nous présentons le mode utilisé pour générer les données et nous introduisons les défauts à étudier. Enfin, nous décrivons les technologies et les outils auxquels nous avons recouru pour aboutir aux résultats présentés au chapitre 3.

### 2.2 Conception du système (Architecture matérielle)

#### 2.2.1 Présentation du système PV étudié

Le système PV que nous étudions est un système PV hors réseau, conçu de manière à reproduire une configuration proche de celle utilisée au niveau du champ PV de la faculté des sciences et technologies (FST) de notre université. Dans notre travail, le système est analysé en circuit ouvert ; ce choix permet de réaliser des simulations représentatives tout en facilitant l'analyse du comportement du système sous différentes conditions de fonctionnement. La Figure 2.1 présente le schéma global du système et ses éléments constitutifs. Nous allons étudier ce système avec la structure suivante :

- Le champ photovoltaïque est constitué de trois chaînes (strings). Chaque chaîne comprend deux groupes de modules photovoltaïques, et chaque groupe est composé de quatre modules connectés en parallèle, chacun ayant une puissance nominale de 319,55 Wc ;
- Une batterie de stockage Lithium-Ion d'une capacité de 45Ah et d'une tension nominale de 51,2V ;

- Un convertisseur DC/DC de type Boost couplé à un dispositif de suivi du point de puissance maximale (MPPT) qui permet d'optimiser la production énergétique du champ PV ;
- Un convertisseur bidirectionnel de type Buck-BOOST qui permet de gérer la charge et la décharge de la batterie ;
- Un système de contrôle intelligent, basé sur des correcteurs PID discrets, permettant d'assurer la stabilité et la régulation du système ;
- Un système d'acquisition des données permettant de recueillir les mesures des grandeurs électriques aux différents points critiques du système PV.

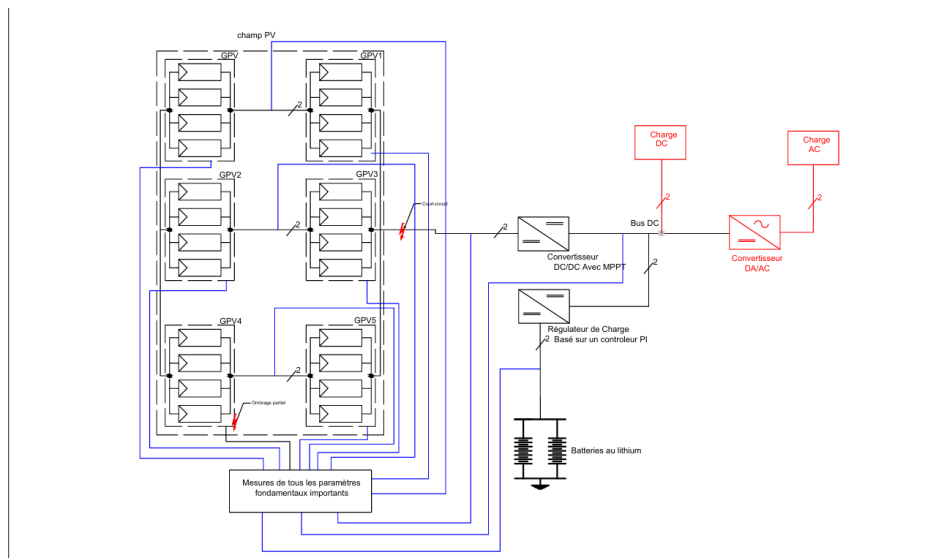


FIGURE 2.1 – Schéma global du système chois

## 2.2.2 Dimensionnement du champ photovoltaïque

### 2.2.2.1 Configuration du champ PV

Comme mentionné précédemment, la structure du champ PV peut être résumée comme suit :

- Nombre total de modules PV ou PV Array : 6
- PV Array par string : 2
- Nombre de strings en parallèle : 3

— Nombre total de modules PV : 24

Cette configuration série-parallèle permet d'adapter les niveaux de tension et de courant du générateur photovoltaïque aux exigences du système. La connexion en série des PV Array permet d'augmenter la tension de sortie, tandis que la connexion en parallèle des strings permet d'augmenter le courant total fourni par le champ PV.

### 2.2.2.2 Calcul de la puissance totale installée

En considérant les descriptions de la section précédente, sachant que chaque module a une puissance nominale de 319,55 Wc dans les conditions standard d'essai, la puissance fournie par un PV Array est donnée par la formule (2.1) et la puissance totale installée par la formule (2.2).

$$\text{Puissance d'un PV array} = \text{nombre des modules} \times \text{puissance d'un module} \quad (2.1)$$

$$P_{total} = \text{Nombre des PV Array} \times P_{array} \quad (2.2)$$

On a alors :  $P_{array} = 4 \times 319,55 = 1278,2 \text{ W} \approx 1,28 \text{ kW}$ .

Le champ PV total est constitué de 6 PV Array, d'où on a :

$$P_{total} = 6 \times 1278,2 = 7669,2 \text{ W} \quad (2.3)$$

La puissance nominale totale du champ photovoltaïque utilisé est donc d'environ 7,67 kWc.

### 2.2.2.3 Justification du dimensionnement

Nous avons dimensionné le champ photovoltaïque de sorte que nous obtenons des niveaux de tension et de courant compatibles avec le fonctionnement du système. L'association de deux PV Array en série permet d'augmenter la tension, tandis que la connexion de trois strings en parallèle d'augmenter le courant. La puissance installée d'environ 7,67 kWc est ainsi suffisante pour alimenter la charge et permettre l'étude des défauts de court-circuit et d'ombrage partiel.

### 2.2.3 Conversion DC/DC (Boost et MPPT)

Le convertisseur DC-DC retenu pour cette étude est de type Boost (élevateur de tension). Ce choix est motivé par le fait que la tension délivrée par le champ photovoltaïque est variable et généralement inférieure à la tension requise par le bus continu commun. Le convertisseur Boost permet ainsi d'élever cette tension tout en assurant une adaptation d'impédance efficace entre la source (panneaux PV) et la charge (bus DC). Sa structure simple, son bon rendement et sa facilité de commande en font une solution parfaitement adaptée aux applications photovoltaïques.

Le fonctionnement repose sur la commutation d'un interrupteur électronique commandé par un signal PWM (Pulse Width Modulation), qui permet de contrôler le rapport cyclique  $D$ . Ce rapport représente le pourcentage de temps pendant lequel l'interrupteur est fermé sur une période de commutation  $T$ . Lorsque l'interrupteur est fermé, l'inductance stocke de l'énergie provenant du champ photovoltaïque ; lorsqu'il s'ouvre, cette énergie accumulée est transférée vers la sortie à travers la diode, permettant d'obtenir une tension supérieure à celle d'entrée.

Afin d'optimiser l'énergie extraite, un algorithme de suivi du point de puissance maximale (MPPT) est implémenté via la méthode *Perturb and Observe* (P&O), choisie pour sa simplicité et son faible coût de calcul. Comme illustré dans l'ordinogramme de la Figure 2.2, la logique de décision, intégrée dans la fonction MATLAB `Dboost`, repose sur deux étapes clés :

- **Observation de la puissance** : Le système calcule la puissance actuelle  $P_{pv} = V_{pv} \times I_{pv}$  et analyse sa variation  $dP_{pv}$  par rapport à l'état précédent stocké en mémoire. Si la puissance augmente ( $dP_{pv} > 0$ ), le système se rapproche du MPP ; si elle diminue ( $dP_{pv} < 0$ ), il s'en éloigne.
- **Ajustement de la perturbation (Action)** : L'algorithme compare la tension actuelle à la précédente ( $dV_{pv}$ ) pour déterminer le sens de la perturbation précédente. Sur la base du produit de ces deux observations ( $dPV_{pv} = dV_{pv} \times dP_{pv}$ ), l'algorithme décide d'augmenter ou de diminuer le rapport cyclique  $D$  par un pas fixe  $D_{step}$ .

Grâce à l'utilisation de variables persistantes (`persistent mem`), le contrôleur ajuste ainsi en continu le rapport cyclique du convertisseur Boost. Cela permet de maintenir le fonc-

tionnement du champ PV au point de puissance maximale en compensant dynamiquement les variations d'irradiance et de température.

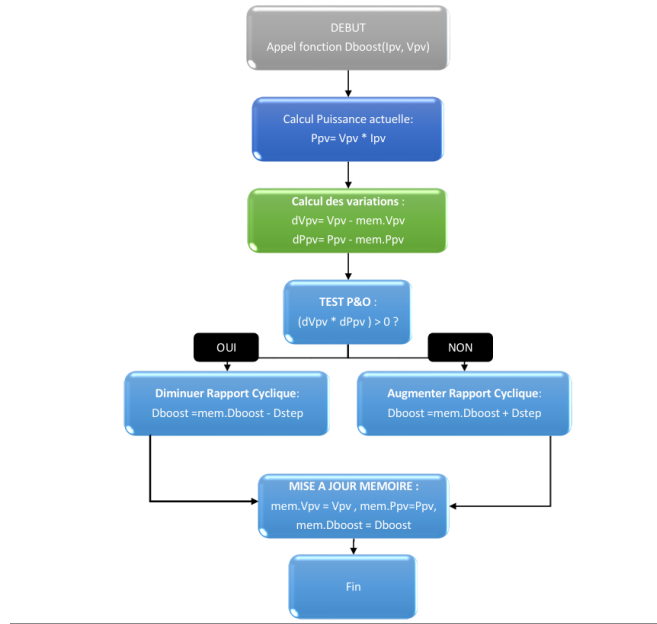


FIGURE 2.2 – Ordinogramme de l’algorithme MPPT (P&O)

## 2.2.4 Gestion du stockage d’énergie

Le stockage de l’énergie (Figure 2.3) est assuré par une batterie Lithium-Ion connectée au bus continu à travers un convertisseur DC/DC bidirectionnel de type Buck-Boost. Ce convertisseur permet de gérer le transfert d’énergie entre le champ photovoltaïque et la batterie.

Lorsque la production photovoltaïque est supérieure à la demande de la charge, l’énergie excédentaire est stockée dans la batterie. Le convertisseur fonctionne alors en mode Buck (charge). Par contre, lorsque la production du champ PV est insuffisante, la batterie fournit l’énergie nécessaire au système et le convertisseur fonctionne en mode Boost (décharge).

Dans la simulation, l’état de la batterie est suivi à l’aide du *State of Charge* (SOC), qui représente le niveau d’énergie disponible dans la batterie et permet d’analyser les phases de charge et de décharge.

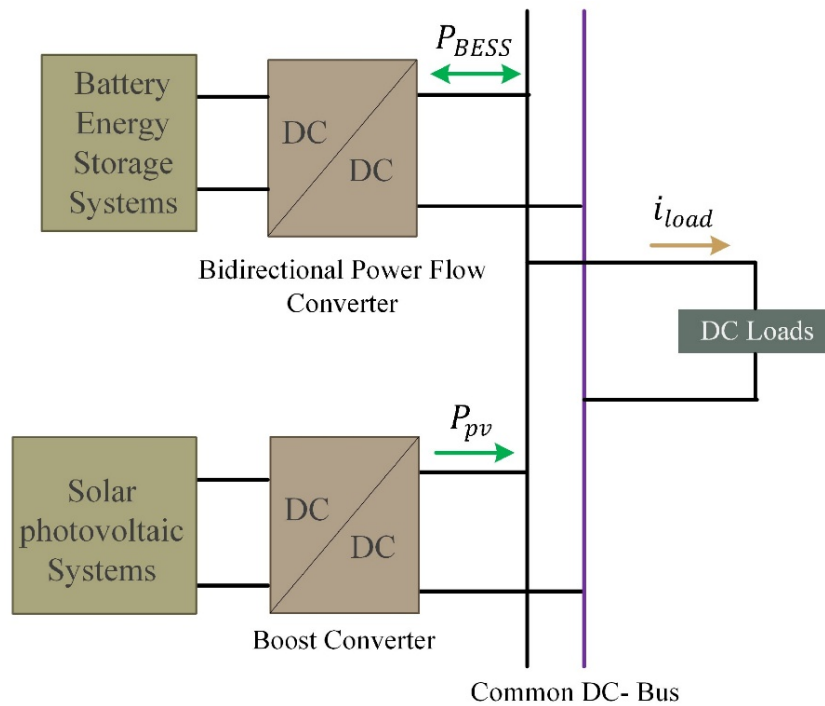


FIGURE 2.3 – Gestion de l'énergie [42]

### 2.2.5 2.2.5 Stratégie de contrôle

La stabilité du système est assurée par une stratégie de contrôle basée sur des boucles de régulation de tension et de courant. Ces boucles permettent de maintenir les grandeurs électriques du système à des valeurs appropriées malgré les variations de la production photovoltaïque et de la charge.

La régulation est réalisée à l'aide de correcteurs PID discrets, implémentés dans l'environnement de simulation. Ces correcteurs ajustent les signaux de commande des convertisseurs afin de garantir une réponse stable et un suivi adéquat des consignes de tension et de courant.

### 2.2.6 2.2.6 Acquisition de données

Le système d'acquisition permet de mesurer les principales grandeurs nécessaires à l'analyse du fonctionnement du système photovoltaïque. Nous utilisons des *scopes* pour visualiser la tension, le courant et l'irradiation solaire, qui influencent directement les performances du champ photovoltaïque.

Les points de mesure sont placés à des endroits stratégiques du système, notamment à la sortie du champ PV, au niveau des modules et au niveau de la batterie, afin de collecter les différentes valeurs du courant, de la tension, de la puissance qui constituent un ensemble de données nécessaires à l'analyse du comportement du système et à la détection des défauts.

Nous utilisons un dispositif d'archivage numérique implémenté via une fonction nommée « `logDataToCSV` » qui nous permet d'exporter les données simulées vers un fichier externe sous format CSV. Les données sont enregistrées de manière temporelle avec un pas d'échantillonnage de 1 microseconde afin de garantir une précision dans la capture des différents phénomènes transitoires.

## 2.3 Détection d'anomalies basée sur l'intelligence artificielle

Nous exploitons l'intelligence artificielle pour détecter automatiquement les anomalies dans le système PV, pour ce faire nous utilisons des techniques de Machine Learning (ML) qui nous aident à analyser les données que nous allons générer lors de la simulation du système PV et à identifier les différents états de fonctionnement de notre champ PV.

### 2.3.1 Identification des anomalies

Les systèmes PV peuvent être affectés par plusieurs types d'anomalies qui réduisent leur rendement. Nous allons nous intéresser à certaines anomalies du champ PV qui sont facilement simulables, principalement l'ombrage partiel et le court-circuit.

On parle d'ombrage partiel lorsqu'une partie du champ PV est partiellement couverte par un obstacle qui empêche au module de recevoir un éclairage maximal, ce qui entraîne une diminution de la puissance produite. Le court-circuit est vu comme un défaut électrique qui provoque une circulation anormale du courant dans le système.

### 2.3.2 Approche utilisée

Pour diagnostiquer l'état de fonctionnement du système photovoltaïque, nous adoptons une approche structurée en quatre étapes principales, allant de la collecte des données à l'évaluation des modèles de classification. Cette méthodologie garantit à la fois rigueur et robustesse, comme illustré par l'ordinogramme global de la Figure 2.4.

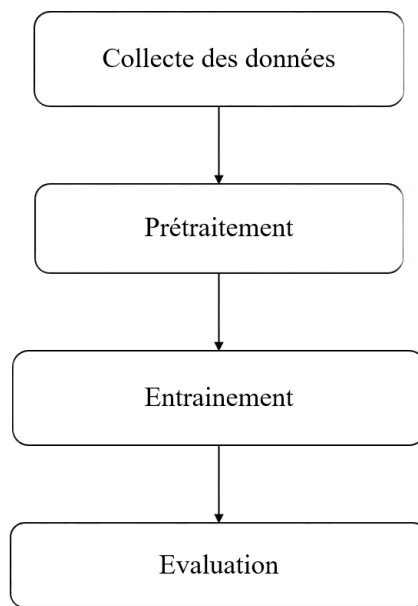


FIGURE 2.4 – Ordinogramme d'apprentissage

#### 2.3.2.1 Collecte et structuration des données

Les variables d'entrée ( $x$ ) au modèle incluent la tension, le courant et la puissance du système PV, qui sont les variables d'entrée souvent considérées dans la littérature. Chaque observation est associée à une étiquette ( $y$ ) (qui constituera la sortie du modèle) représentant l'état réel du système : fonctionnement normal, ombrage partiel ou court-circuit. L'ensemble des couples  $(x, y)$  constitue la base étiquetée que l'on utilise pour l'apprentissage supervisé; ce processus appelé « classification » permet donc d'associer un groupe de valeurs électriques à un état du système et entraîner le modèle à prédire la classe correspondante.

### 2.3.2.2 Prétraitement des données

Pour améliorer la qualité des données issues de la simulation, on passe d'abord à un prétraitement de ces dernières avant d'entraîner les modèles de ML, ainsi les algorithmes n'auront pas du mal à les exploiter.

Le prétraitement commence par la préparation des données contenues dans le dataset issus de la simulation que nous labellisons en faisant correspondre chaque classe à un numéro : le fonctionnement normal, l'ombrage partiel et le court-circuit respectivement associés à 0, 1 et 2. Ensuite, nous divisons les données en deux ensembles dont l'un est destiné à l'entraînement du modèle et l'autre à la prédiction afin d'évaluer ses performances.

Nous poursuivons avec la normalisation des données qui a pour objectif de ramener les différentes valeurs à une même échelle, cette étape est surtout importante pour certains algorithmes de ML en particulier pour les SVM qui sont sensibles aux différences d'échelle entre variables.

Enfin, une réduction de dimension peut être faite pour simplifier l'ensemble des données sans perdre l'information essentielle. Après ces étapes, nous utilisons les données pour l'entraînement des modèles suivi de l'évaluation de leurs performances.

### 2.3.2.3 Entraînement et pipeline des modèles

Chaque algorithme est entraîné sur l'ensemble d'apprentissage afin de prédire les trois états du système. Les paramètres de chaque modèle jouent un rôle critique, influençant directement la précision, la robustesse et la capacité de généralisation.

**Machine à vecteurs de support (SVM)** Le SVM maximise la marge entre les classes et utilise différents noyaux pour gérer la non-linéarité des données. Les paramètres principaux sont :

- **C (coefficient de régularisation)** : équilibre entre la maximisation de la marge et la tolérance aux erreurs sur l'ensemble d'apprentissage.
- **$\gamma$  (paramètre du noyau)** : contrôle la forme de la frontière de séparation, influençant la capacité du modèle à capturer les zones d'anomalies.
- **Sortie** : prédiction de l'état du système.

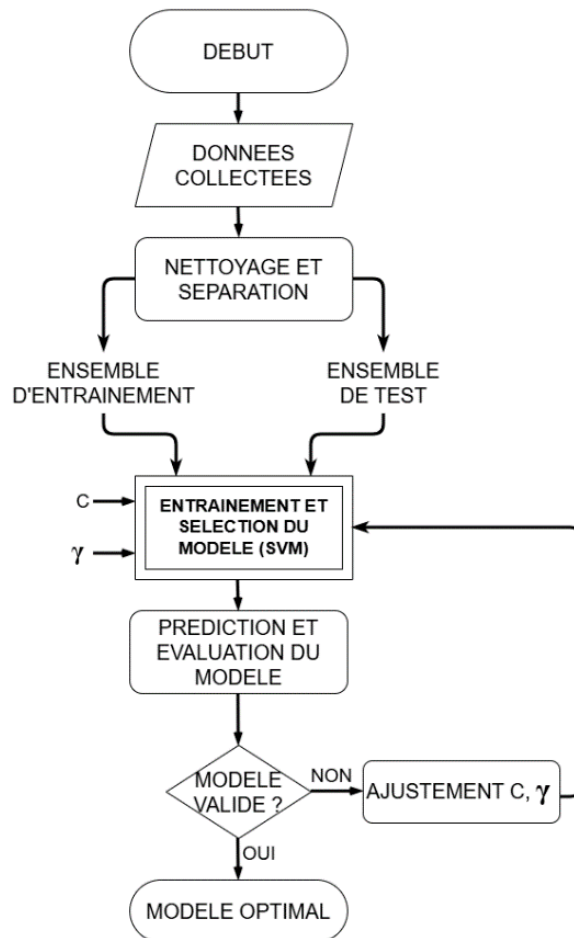


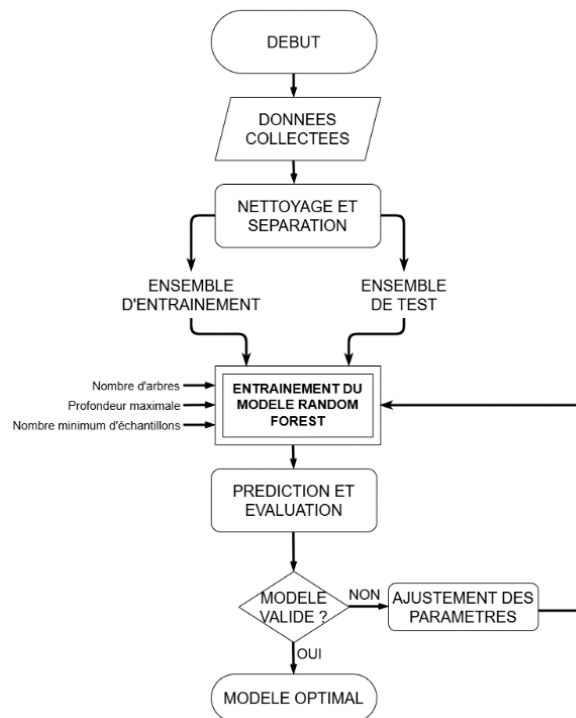
FIGURE 2.5 – Flow chart d’entraînement et de prédiction du SVM [43]

La Figure 2.5 illustre le processus de diagnostic par Machine Learning qui débute par la collecte des données brutes, leur nettoyage, leur normalisation et leur division. L’étape centrale est l’entraînement du modèle SVM pour apprendre à distinguer les signatures électriques. Enfin, le modèle est déployé sur l’ensemble de test pour classer l’état du système.

**Random Forest (Forêt aléatoire)** La Random Forest est adaptée aux relations complexes entre variables et réduit le risque de surapprentissage grâce à la combinaison d’arbres de décision. Paramètres principaux :

- **Nombre d’arbres** : plus le nombre est élevé, plus le modèle est robuste, mais le temps de calcul augmente.

- **Profondeur maximale** : limite la complexité des arbres pour éviter le surapprentissage.
- **Nombre minimum d'échantillons par feuille** : contrôle la granularité des décisions et la capacité à généraliser.
- **Sortie** : prédiction de l'état du système.



]Flowchart du Random Forest [44]

Le processus de fonctionnement reste le même que pour le SVM, seuls les paramètres changent comme on peut le voir sur la Figure 2.3.2.3.

**Arbre de décision (Decision Tree)** L'arbre de décision offre une approche interprétable et rapide. Chaque nœud applique une condition sur une variable électrique :

- **Critère d'impureté (Gini)** : mesure l'homogénéité des classes à chaque nœud.
- **Profondeur maximale** : empêche l'arbre de devenir trop complexe.
- **Taille minimale des feuilles** : assure que chaque décision repose sur un nombre suffisant d'observations.

- **Sortie** : prédiction de l'état du système.

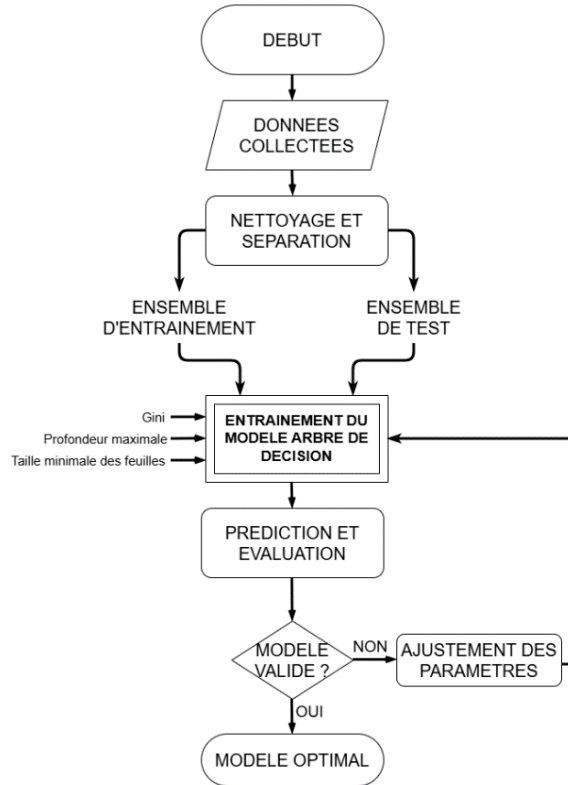


FIGURE 2.7 – Flowchart de l'arbre de décision

La figure 2.7 illustre le processus de fonctionnement, qui reste identique au SVM à l'exception des paramètres spécifiques.

### 2.3.2.4 Évaluation des modèles

L'évaluation repose sur l'ensemble de test, jamais utilisé lors de l'entraînement, pour calculer :

- Précision (Accuracy)
- Rappel (Recall)
- Score F1

Ces métriques permettent de comparer les modèles, de valider leur capacité à détecter les ombres partiels et les court-circuit, et de justifier le choix des paramètres pour chaque algorithme.

## 2.4 Technologies et outils utilisés

Pour simuler le système présenté à la section 2.2 et entraîner les modèles de la section 2.3, nous utilisons différents outils tels que :

### 2.4.1 MATLAB / Simulink

MATLAB (Matrix Laboratory), développé par Math Works, est un environnement de calcul numérique et un langage de programmation de haut niveau plébiscité par la communauté scientifique et technique. Sa force réside dans son architecture interactive qui permet une manipulation native et directe des structures matricielles et vectorielles, le distinguant ainsi des langages conventionnels comme le C, le Java ou le Fortran.

Cet écosystème intègre de manière fluide le calcul mathématique, la programmation et la visualisation graphique au sein d'une interface intuitive. Grâce à sa polyvalence, il couvre un large spectre d'applications allant de l'algèbre linéaire et l'optimisation au traitement du signal et à l'analyse de données. En outre, MATLAB facilite l'implémentation d'algorithmes complexes et offre une grande interopérabilité grâce à sa capacité d'interfaçage avec d'autres langages tels que le C++, Python et Java.

Le logiciel propose également des bibliothèques spécialisées, ou *toolboxes*, qui étendent ses fonctionnalités à des domaines de pointe comme l'apprentissage automatique et le contrôle des systèmes. Dans le cadre de notre étude, nous allons notamment exploiter la *Signal Processing Toolboxes* pour le traitement et l'analyse des grandeurs électriques.

Par ailleurs, Simulink, extension incontournable de MATLAB, a joué un rôle central dans la conception et la simulation de notre système électrique. Grâce à son interface graphique intuitive basée sur des schémas blocs, Simulink a permis une modélisation précise des composants photovoltaïques, la réalisation de simulations dynamiques et l'extraction rigoureuse des données nécessaires à l'analyse des défauts.

La Figure 2.8 montre à quoi ressemble l'interface Simulink de MATLAB avec ses éléments, entre autres :

— **La Barre d'Outils et de Menu (Haut) :**

— *Menu File/Edit/View* : Permet de gérer le fichier de modèle.

- *Display* : Le bouton vert permet de lancer la simulation temporelle du système.
- *Diagram* : permet d'effectuer des modifications personnalisées sur les blocs utilisés dans le modèle (rotation, format, etc.).
- *Simulation* : cette zone nous permet de configurer les paramètres et d'exécuter le modèle.
- *Analysis* : permet d'extraire les données de performance.
- *Code* : permet de générer de code (C/C++) à partir des blocs.
- *Tools* : permet d'accéder aux différentes bibliothèques et d'explorer le modèle.
- *Help* : permet d'accéder aux documentations par rapport au contenu des éléments.
- **La zone de Modélisation** : C'est l'espace où on dispose et connecte les blocs reliés par des signaux afin de former un modèle prêt à la simulation.
- **Barre d'État (Bas)** : Elle donne des informations sur l'état actuel du logiciel.
  - *Ready / Running* : Indique si le modèle est prêt ou si le calcul est en cours.
  - *Pourcentage* : Indique l'avancement de la simulation.
  - *VariableStepAuto* : Indique que Simulink ajuste automatiquement le pas de calcul pour optimiser la précision et la vitesse.

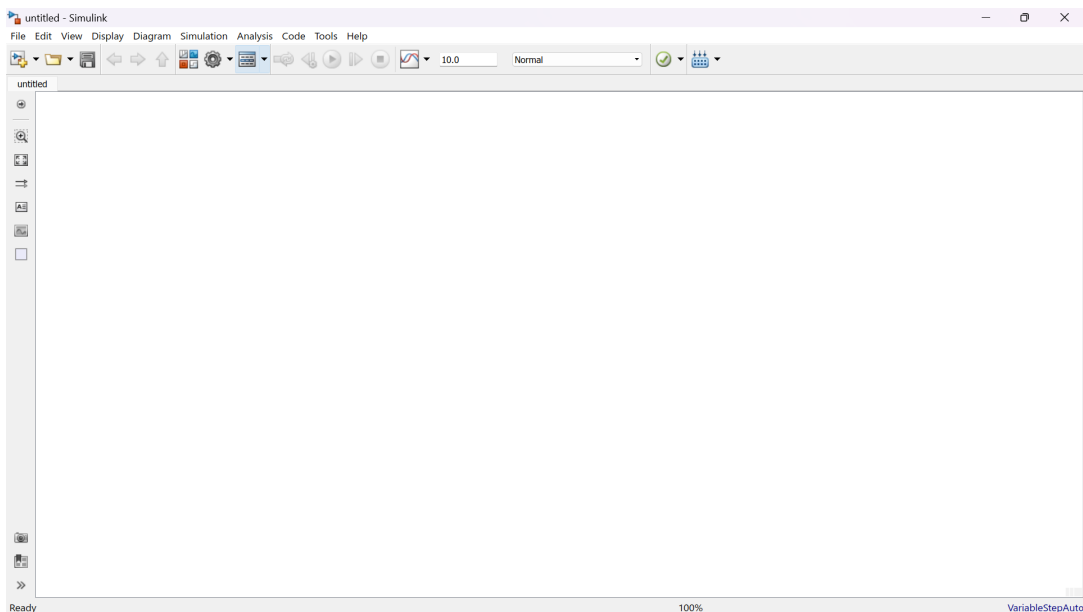


FIGURE 2.8 – Interface de Simulink

## 2.4.2 Jupyter notebook

Nous utilisons Jupyter Notebook pour nous permettre d'écrire et d'exécuter du code Python tout en visualisant directement les résultats. L'interface de Jupyter est composée principalement de :

- **La barre de menu (Haut)** : Elle se situe en haut de l'interface et elle regroupe les options comme File, Edit, View, Run et Kernel, qui nous permettent de gérer les fichiers, modifier le contenu du notebook, d'exécuter le code et /ou contrôler le noyau d'exécution.
- **La barre d'outil (Haut)** : Située juste en bas de la barre de menu, elle contient les icônes qui facilitent certaines opérations courantes comme la sauvegarde du notebook, l'exécution des cellules de code ou l'ajout de nouvelles cellules.
- **La zone de saisie (Centre)** : C'est la partie centrale de l'interface qui est constituée des cellules, qui sont les espaces où l'utilisateur écrit du code Python ou du texte explicatif. Chaque cellule peut être exécutée indépendamment, ce qui nous permet d'analyser les résultats de manière progressive.
- **Noyau (Kernel) Python (Haut)** : Il se situe à droite de la barre d'outils et indique l'environnement d'exécution utilisé pour interpréter le code.

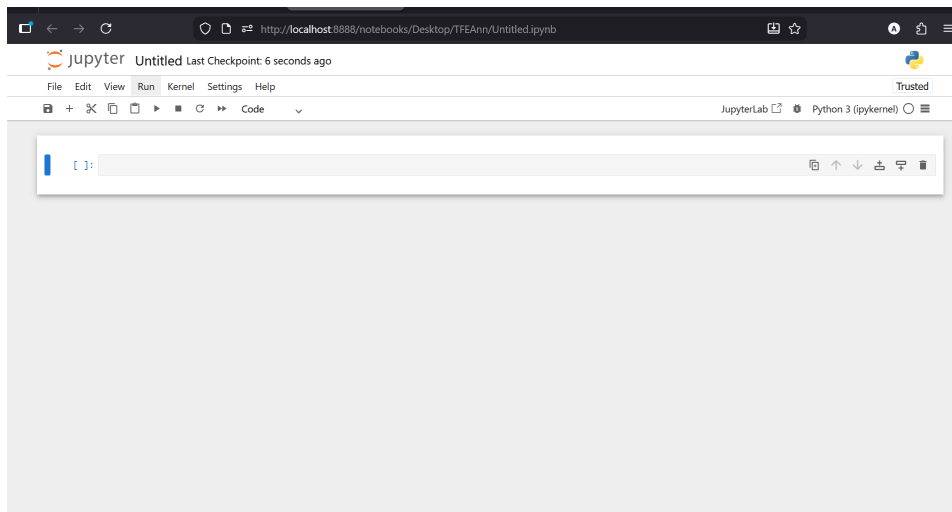


FIGURE 2.9 – Environnement Jupyter notebook

Pour parvenir à implémenter les modèles que nous étudions, nous avons utilisé plusieurs

bibliothèques de Python via l'environnement Jupyter notebook (Figure 2.9) parmi lesquelles nous avons « NumPy » qui nous facilite à faire des calculs, « Pandas » qui est utilisé pour gérer et analyser les données, « Scikit-learn » qui nous permet d'implémenter les algorithmes de ML comme par exemple ceux que nous utilisons (SVM, Random Forest et les arbres de décision), « Matplotlib » et « Seaborn » qui nous ont servi pour la visualisation des résultats obtenus.

## 2.5 Conclusion partielle

Dans cette partie de notre travail, nous parlons de la modélisation, de la configuration du système PV, des architectures des différents modèles que nous utilisons dans la suite du travail ainsi que des outils utilisés.

Avec ce qui précède, dans le chapitre suivant nous abordons un modèle simulé d'un système PV qui nous permet de générer les données utiles au développement et à l'évaluation des algorithmes de détection.

# Chapitre 3

## Implémentation et simulation du modèle d'un système PV pour la génération des données

---

### 3.1 Introduction

Dans ce chapitre, nous présentons la simulation du système photovoltaïque étudié ainsi que la méthode adoptée pour la gestion des données issues de cette simulation.

### 3.2 Présentation du modèle du système PV

Nous nous sommes appuyés sur le schéma présenté à la Figure 2.1 pour implémenter, sous Simulink, le modèle du système photovoltaïque illustré à la Figure 13. Celui-ci sera étudié en conditions de circuit ouvert, des défauts de court-circuit ainsi que sous irradiation variable, afin de simuler son fonctionnement en régime normal et en présence de défauts, et de générer les données nécessaires à l'analyse. L'objectif principal de notre travail étant la collecte des données, nous omettons certains détails spécifiques par rapport au choix des paramètres utilisés lors de la modélisation du système pendant la simulation.

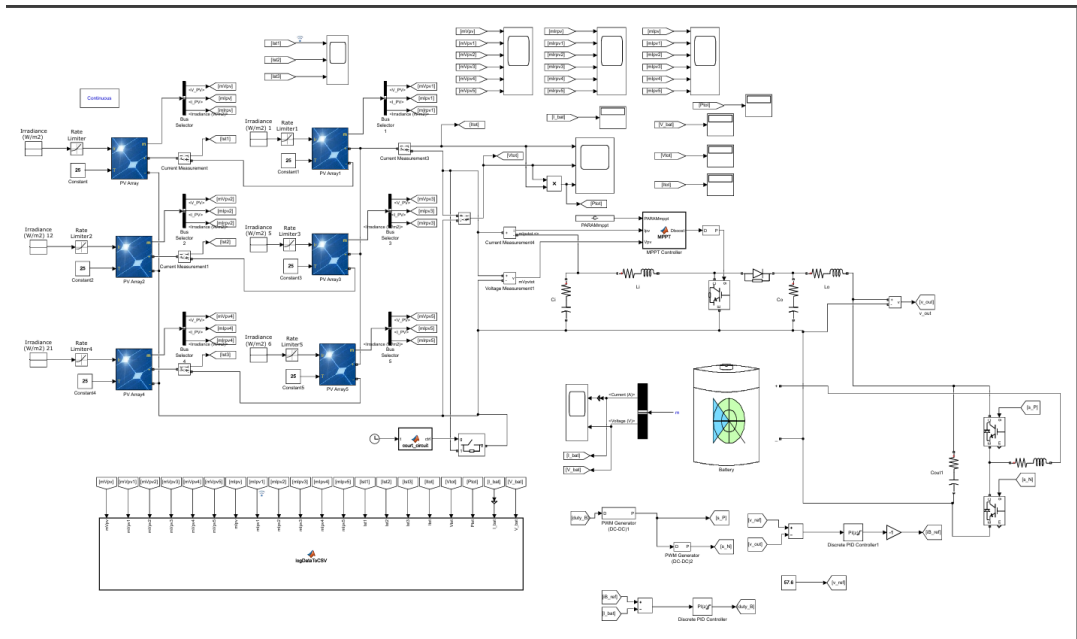


FIGURE 3.1 – Modèle Simulink du système choisi

## 3.3 Implémentation du modèle du système PV

### 3.3.1 Implémentation du champ PV

Nous avons utilisé le bloc PV Array de la bibliothèque Simscape, dont les caractéristiques sont présentées à la Figure 3.2(a). Le champ photovoltaïque (Figure 3.2(b)) est composé de trois strings en parallèle, chaque string comprenant deux modules en série, et chaque module quatre sous-modules en parallèle. L'ensemble constitue 24 modules pour une puissance crête de 7,67 kWc sous une tension de 54,48V au point de puissance maximale (PPM).

Des capteurs de courant mesurent le courant de chaque string et sont reliés aux blocs d'affichage et d'acquisition de données. Des bus de mesure permettent également de récupérer les tensions, courants et irradiations des modules, qui sont ensuite transmis au bloc d'acquisition.

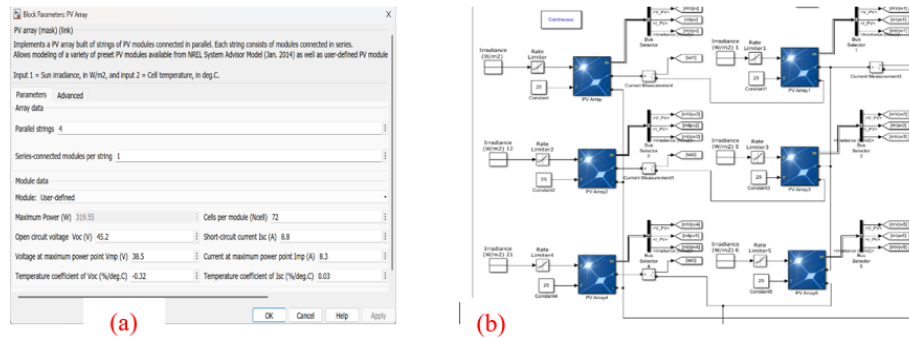


FIGURE 3.2 – (a) Caractéristiques du bloc PV, (b) Champ PV

### 3.3.2 Implémentation du convertisseur DC/DC

Cette étape est représentée à la Figure 3.3. Elle est faite par des blocs de Simulink et se compose d' :

- Une fonction Matlab qui implémente le script de l'algorithme P&O pour le suivi du PPM;
- Un hacheur Boost composé des résistances, condensateurs, inductances, d'une diode et d'un transistor bipolaire à grille isolée (IGBT), piloté par un algorithme P&O.

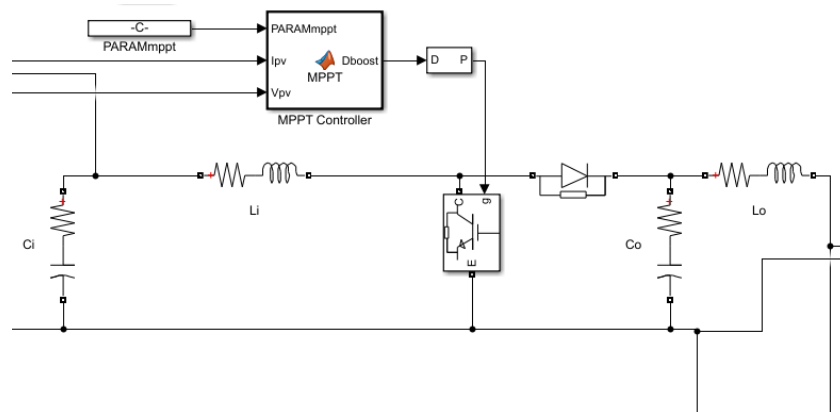


FIGURE 3.3 – Convertisseur DC/DC

### 3.3.3 Modèle de la batterie de stockage

La Figure 3.4(a) présente le modèle de la batterie de stockage au Lithium-Ion de 51,2 V et 45 Ah, initialement chargée à 50 % ; nous y trouvons un port de mesure permettant de suivre l'évolution du courant et de la tension au travers d'un bus de données.

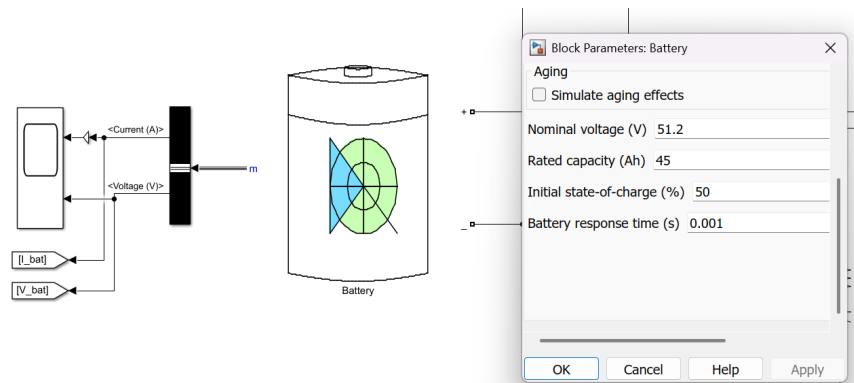


FIGURE 3.4 – Batterie de stockage (a) et ses caractéristiques (b)

### 3.3.4 Convertisseur DC/DC bidirectionnel (type Buck-Boost)

La Figure 3.5 montre le convertisseur DC/DC bidirectionnel que nous avons utilisé. Il nous permet de contrôler l'énergie au sein du système : en mode buck, lorsque les modules PV produisent suffisamment d'énergie, cette dernière abaisse la tension et limite le courant qui circule vers la batterie pour éviter la surcharge. En mode boost, lorsque la production est faible, il élève la tension et permet à la batterie de renvoyer son énergie vers le bus DC du système.

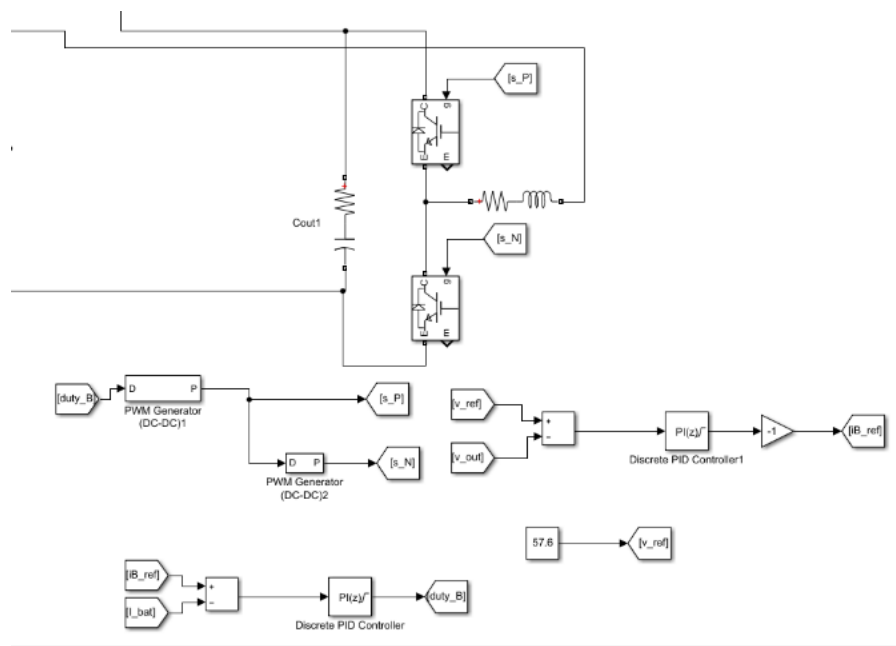


FIGURE 3.5 – Convertisseur bidirectionnel Buck-Boost

### 3.3.5 Implémentation du système d'acquisition des données électriques

Le module d'acquisition des données électriques de la Figure 3.6 a été obtenu en utilisant un bloc fonction Matlab avec le script (Annexe A) et sert à générer directement un dataset sous format csv exploitable pour le ML.

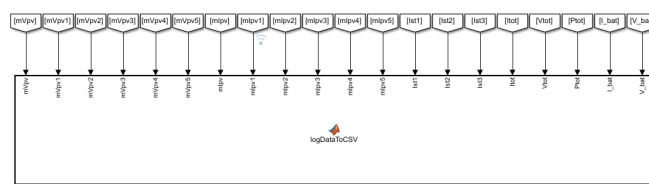


FIGURE 3.6 – Bloc d'acquisition des données

## 3.4 Simulation du système PV et génération des données

### 3.4.1 Etiquetage des grandeurs mesurées

Les marques de la Figure 3.7 indiquent les points où nous avons introduit les défauts. Nous y trouvons également les 20 grandeurs mesurées et acquises entre autres :

#### Au niveau du champ PV

- $mV_{pv}$ ,  $mV_{pv1}$ ,  $mV_{pv2}$ ,  $mV_{pv3}$ ,  $mV_{pv4}$ ,  $mV_{pv5}$  : respectivement tension aux bornes des modules PV, PV1, PV2, PV3, PV4 et PV5 du champ PV.
- $mI_{pv}$ ,  $mI_{pv1}$ ,  $mI_{pv2}$ ,  $mI_{pv3}$ ,  $mI_{pv4}$ ,  $mI_{pv5}$  : courants respectifs générés par les 6 modules du champ PV.
- $I_{st1}$ ,  $I_{st2}$ ,  $I_{st3}$  : courant respectifs dans les strings (chaines) 1, 2 et 3.
- $V_{tot}$ ,  $I_{tot}$ ,  $P_{tot}$  : la tension, le courant et la puissance totale à la sortie du champ PV.

### Au niveau de la batterie

- $V_{bat}$  : tension aux bornes de la batterie ;
- $I_{bat}$  : courant généré ou absorbée par la batterie.

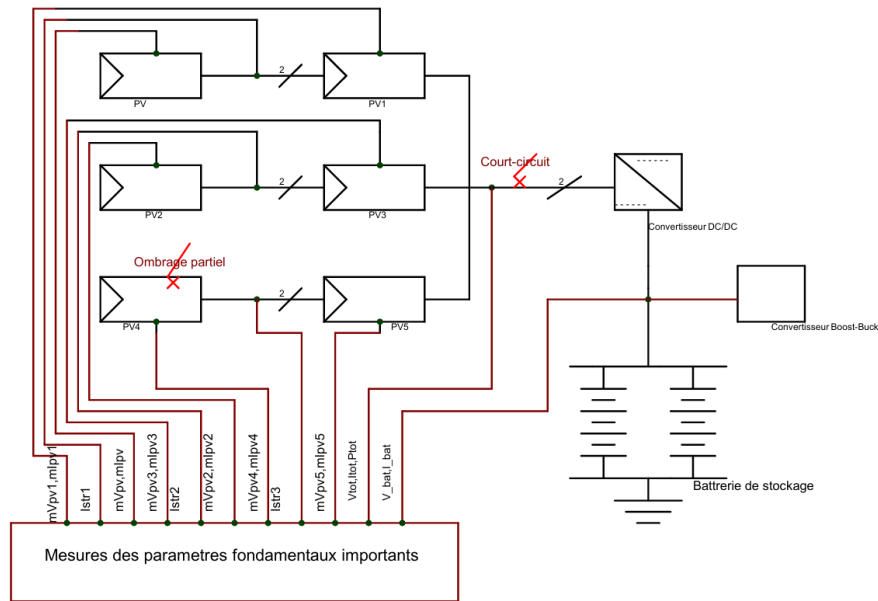


FIGURE 3.7 – Etiquetage des grandeurs et marques des défauts

### 3.4.2 Simulations des états du système

Nous fixons des hypothèses de simulation que voici :

- La température du champ PV reste constante durant toutes les simulations.
- L'irradiance reste constante au cours de la simulation (fonctionnement normal et en court-circuit) et ne change que dans le cas d'ombrage partiel ou on varie l'irradiation du module PV4.
- Après le défaut, le système revient en fonctionnement normal.
- Nous négligeons l'influence des câbles qui relient différents éléments.

Tenant compte des hypothèses qui précèdent, nous étudions et interprétons les différents états de fonctionnement du système PV :

### 3.4.2.1 Fonctionnement normal

Cette simulation dure 3 s (secondes) et se fait sous une irradiation de  $1000W/m^2$  et une température de  $25^{\circ}C$ , le temps d'échantillonnage est de  $1\mu s$  pour l'acquisition des données. L'irradiation est constante pendant toute la simulation comme nous indique la Figure 3.8 (hypothèse de simulation).

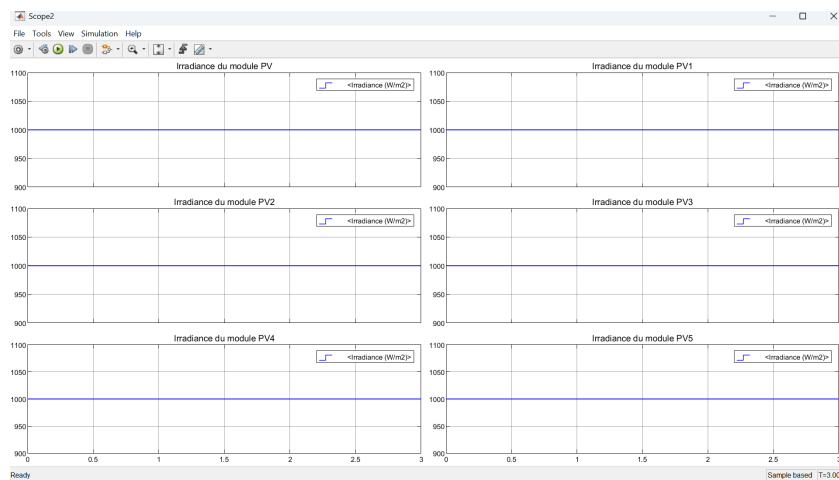


FIGURE 3.8 – Irradiations des modules PV en fonction du temps

La Figure 3.9 montre les différentes courbes des courants générés par les modules PV durant toute la simulation mais on observe des petites oscillations au début de la simulation ( $t < 0,2s$ ), ce qui correspond à la phase de démarrage du système.

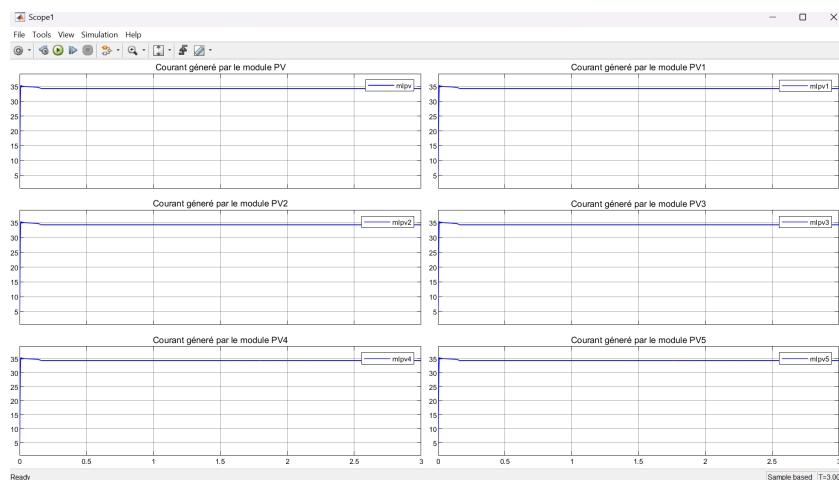


FIGURE 3.9 – Variations des courants générés par les modules PV en fonction du temps

La Figure 3.10 montre que les modules atteignent la valeur nominale vers 0,2s; ce qui confirme que le système atteint rapidement un régime permanent. Comme les modules reçoivent la même irradiation et fonctionnent dans les mêmes conditions, leurs tensions restent identiques durant toute la simulation.

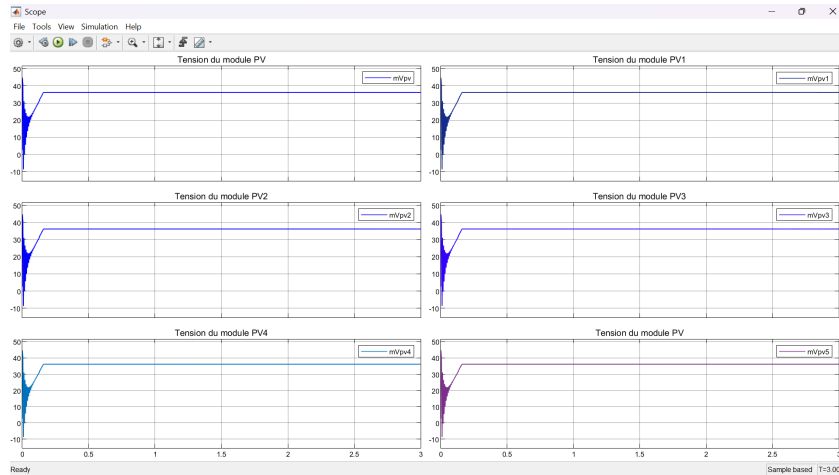


FIGURE 3.10 – Tensions aux bornes des modules PV

Au démarrage, le courant, la tension et la puissance augmentent progressivement jusqu'à atteindre une valeur stable qui correspond au point de puissance maximale du champ PV (Figure 3.11). Après 0,2 s, la puissance produite devient constante car les conditions d'irradiation et de température restent inchangées.

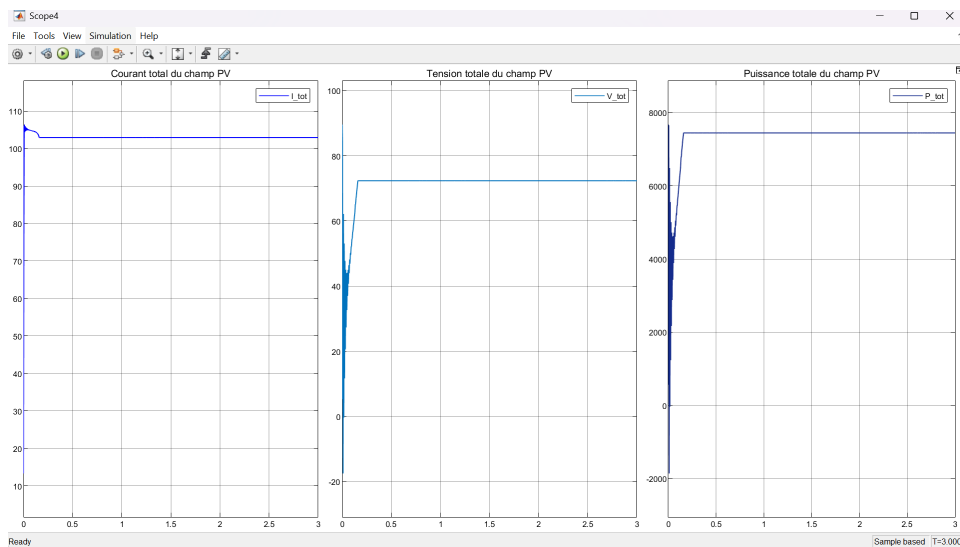


FIGURE 3.11 – Variations des courbes du courant généré, de la tension aux bornes du champ PV et de la puissance fournie par le champ PV

A la Figure 3.12 on observe une certaine instabilité de la tension et le courant de la batterie au départ due à la phase de démarrage du système. Au bout de 0,2 s les deux grandeurs se stabilisent, la batterie fonctionne dans son régime normal.

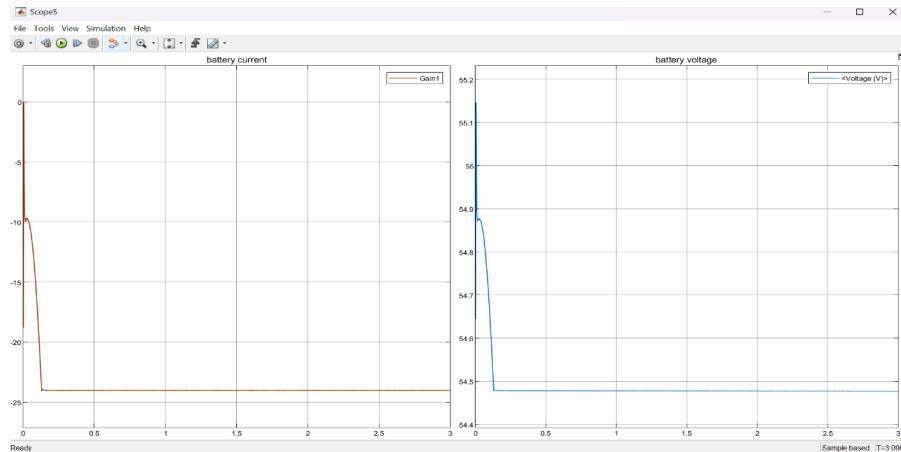


FIGURE 3.12 – Variations des courbes du courant généré ou absorbée par la batterie et la tension aux bornes de la batterie

A la fin de la simulation les données de fonctionnement sont stockées directement dans un fichier nommé « Fonctionnement Normal » sous format .csv.

### 3.4.2.2 Défaut d'ombrage partiel

L'ombrage partiel est appliqué sur le module PV4, cela en variant progressivement l'irradiation grâce à une matrice selon le temps :

- $0s \leq t \leq 0,5s$  :  $1000W/m^2$  (fonctionnement normal)
- $0,5s \leq t \leq 1s$  :  $750W/m^2$
- $1s \leq t \leq 1,5s$  :  $500W/m^2$
- $1,5s \leq t \leq 2s$  :  $750W/m^2$
- $2s \leq t \leq 3s$  : on retourne à  $1000W/m^2$

Les autres modules restent dans les conditions normales (Figure 3.13) de  $1000W/m^2$  (hypothèse).

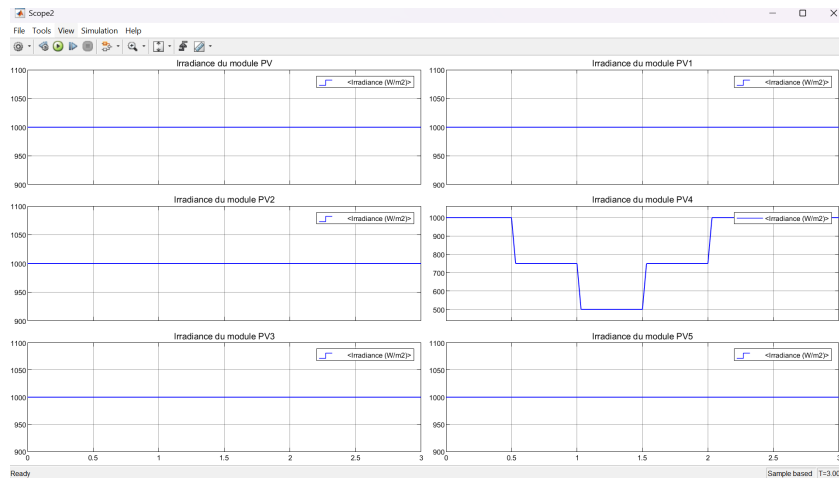


FIGURE 3.13 – Ombrage partiel au niveau du module PV4

A partir de la Figure 3.14, on observe des variations de tensions sur le module PV5 car il se trouve sur le même string que le module ombré (PV4). Entre 0,5 s et 2 s, la tension du module PV4 varie à cause de la réduction de l'irradiation. Lorsque l'ensoleillement diminue, la cellule produit moins de courant. Pour maintenir l'équilibre électrique dans la chaîne de modules montés en série, les tensions des modules voisins s'ajustent. Cette compensation explique les variations opposées observées entre PV4 et PV5, permettant de maintenir relativement stable la tension totale du champ PV.

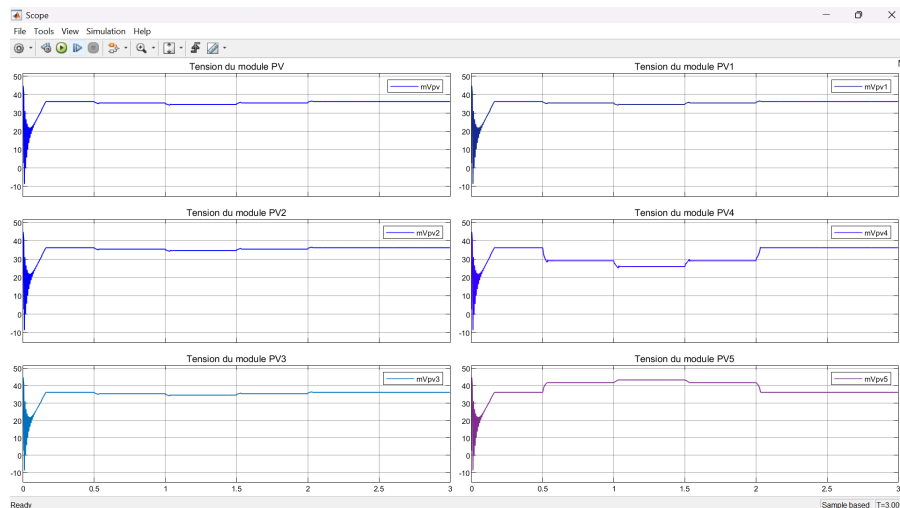


FIGURE 3.14 – Variations des tensions aux bornes des modules avec PV ombrée

Entre 0,5 s et 2 s, les courants de PV4 et PV5 subissent des variations à cause d'une baisse d'ensoleillement (Figure 3.15), cette variation s'explique par le fait que le courant

d'un module PV est directement proportionnel à l'irradiation qu'il reçoit. Ainsi, lorsque l'irradiation diminue, le courant généré diminue également. Les courants des modules PV4 et PV5 remontent à leur valeur normale lorsque l'irradiation revient à  $1000W/m^2$ .

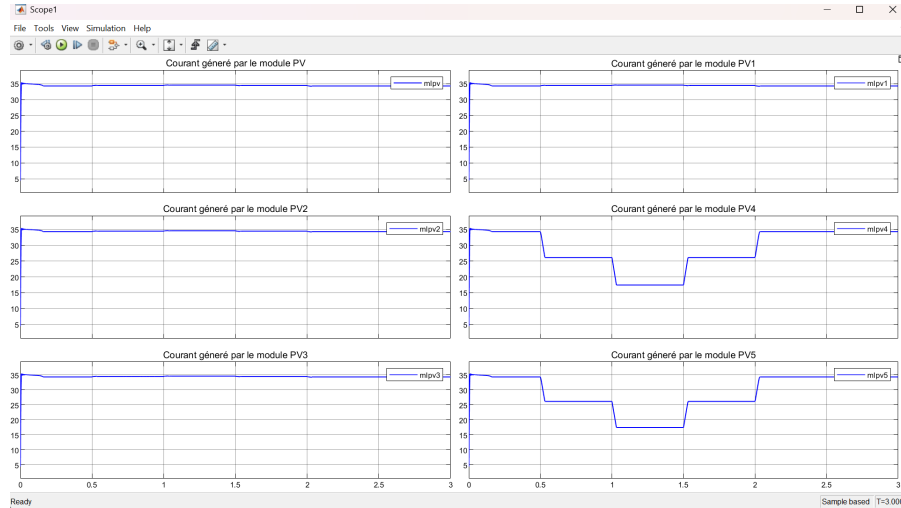


FIGURE 3.15 – Variations des tensions des modules en ombrage partiel

A travers la Figure 3.16 nous remarquons que l'ombrage partiel a agi sur les grandeurs de sortie du système PV. La puissance produite (Figure 3.16(c)) est stable avant  $t = 0,5s$  ensuite, on observe des variations aux instants  $t = 0,5s$ ;  $t = 1s$ ;  $t = 1,5s$  et  $t = 2s$ .

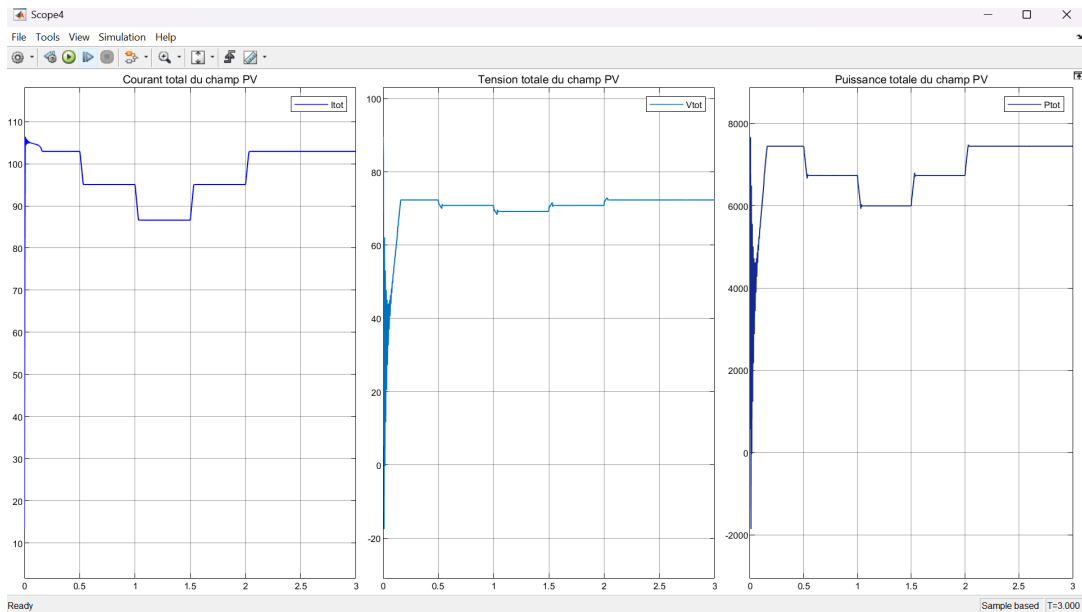


FIGURE 3.16 – Variations du courant (a), de la tension (b) et de la puissance totale (c) durant l'ombrage partiel

Après la phase de démarrage ( $t > 0,2s$ ), le courant et la tension de la batterie restent pratiquement constants, même lorsque l'ombrage partiel est introduit entre 0,5s et 2s (Figure 3.17). Cela montre que la batterie joue un rôle de stabilisation du système.

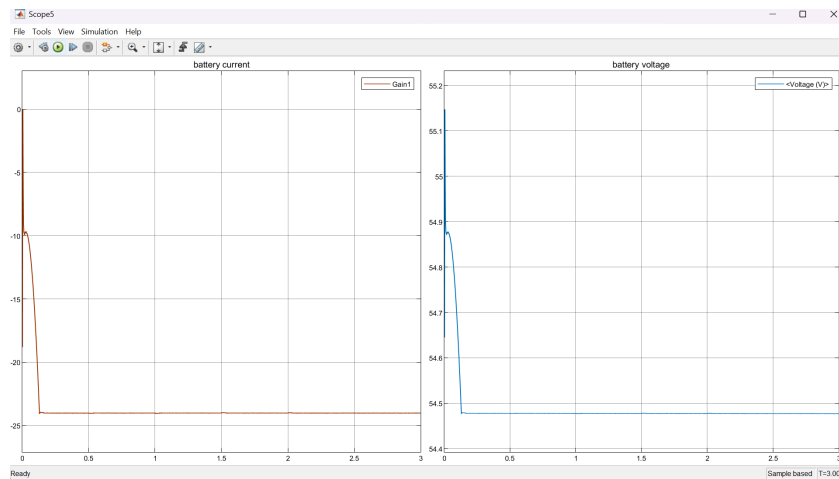


FIGURE 3.17 – Courbes des grandeurs de la batterie de stockage durant l'ombrage

Les données de simulation des défauts d'ombrage sont stockées dans le fichier nommé `Simulation_TfeAn26.csv`.

### 3.4.2.3 C. Défaut de court-circuit

Dans cette étude, le défaut est modélisé par l'activation d'un « Idéal switch » qui court-circuite le champ PV, représentant ainsi un contact accidentel ou une défaillance d'isolation. Nous simulons le défaut de court-circuit pendant 10ms introduite entre 0,5 et 0,51s de la simulation sous une irradiation de  $1000W/m^2$  et une température de 25 degrés Celsius (cfr. Hypothèse).

À l'instant précis du défaut, on observe une chute brutale et totale de la tension pour chaque module comme on peut le voir à la Figure 3.18. Cette chute est due au court-circuit qui crée une résistance presque nulle dans le circuit, ce qui provoque une chute brusque de la tension. Lorsque la durée préétablie du court-circuit s'écoule, la tension se rétablit ramenant ainsi le système à son fonctionnement normal.

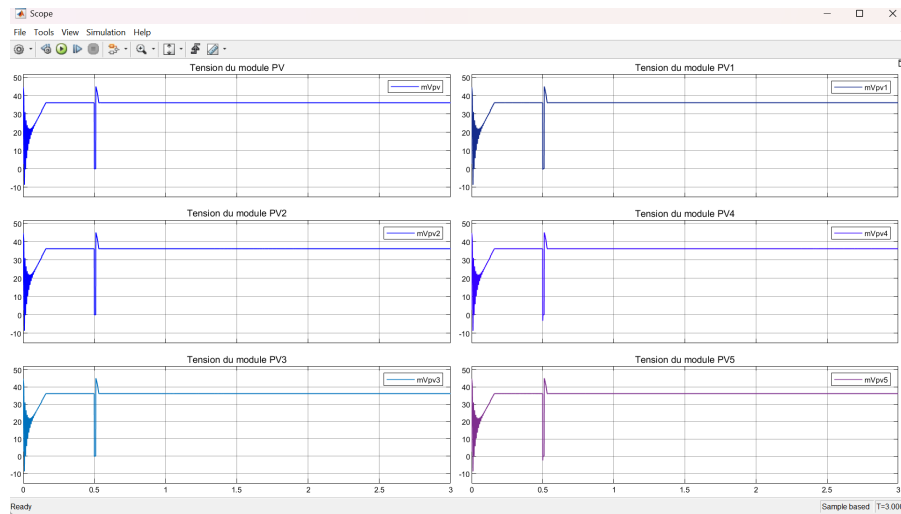


FIGURE 3.18 – Variations des tensions aux bornes des modules PV

Le défaut de court-circuit se manifeste par un pic de courant à l'instant  $t = 0,5s$  (Figure 3.19). Ce phénomène s'explique par la loi d'ohm :

$$I = V/R \quad (3.1)$$

Lorsque la résistance devient faible, le courant augmente : ce qui produit ce pic transitoire, la même chose s'observe sur les strings (Figure 3.20). Après ce pic, le courant diminue et se stabilise quand le système retrouve son état normal.

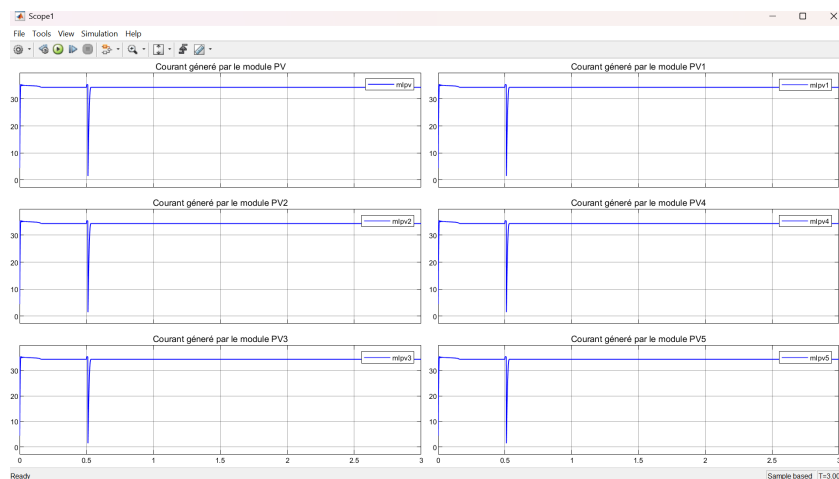


FIGURE 3.19 – Variations des courants générés par les modules PV

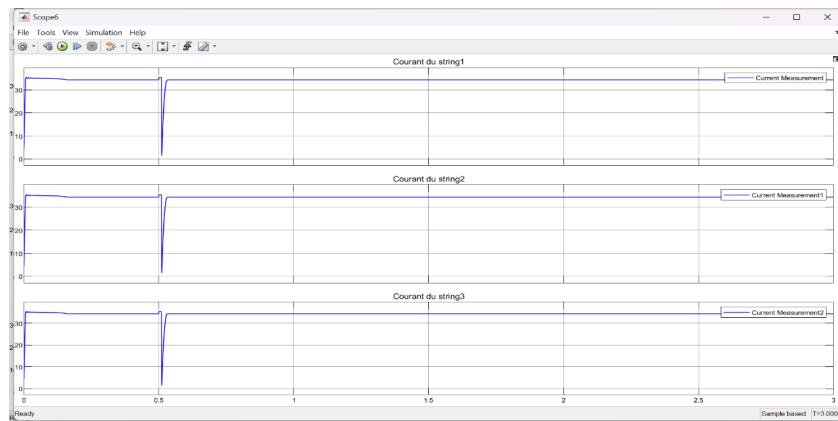


FIGURE 3.20 – Variations des courants dans les strings du champ PV

À exactement 0,5 s ; un court-circuit brutal frappe toute l'installation solaire, faisant tomber la tension (Figure 3.21(b)), la puissance presque à zéro (Figure 3.21(c)) et laissant apparaître un pic transitoire du courant (Figure 3.21(a)). Le système réagit instantanément avec un petit pic de tension et de puissance avant de reprendre son fonctionnement normal (Figure 3.21).

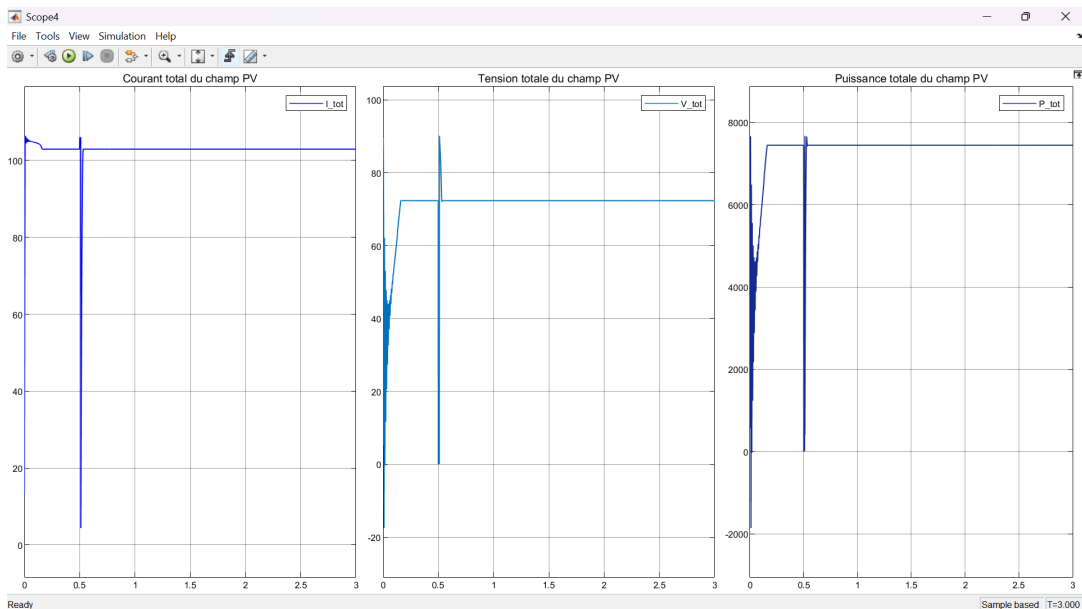


FIGURE 3.21 – Variations de la tension (b), du courant (a) et de la puissance totale (c) du champ PV

La Figure 3.22 démontre qu'au moment du défaut ( $t = 0,5s$ ), la batterie présente : un léger pic de courant et une petite chute de tension. Ces variations sont dues à la

perturbation introduite dans le système par le court-circuit. Après cet instant, la batterie retrouve son régime stable.

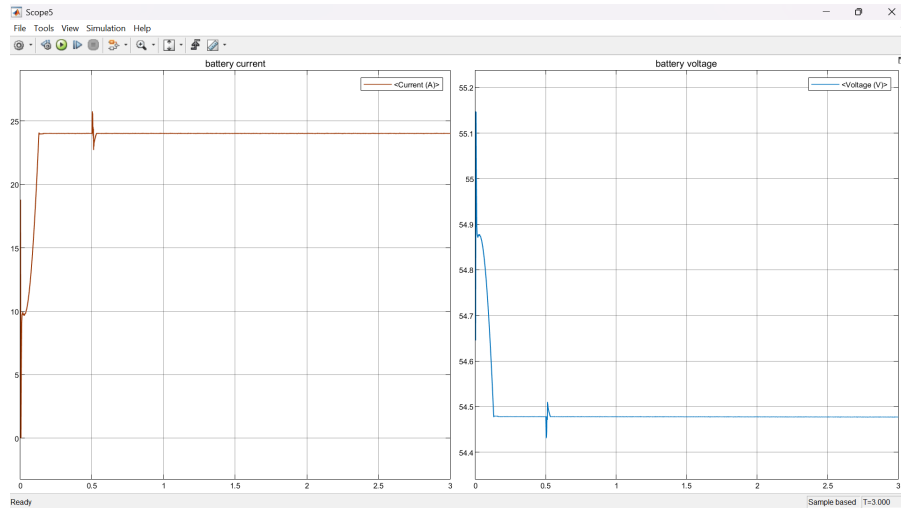


FIGURE 3.22 – Variations des tensions de la batterie (a) et courant généré par la batterie (b)

### 3.5 Conclusion partielle

Dans cette partie du travail, nous avons simulé le cas type du système similaire à celui de notre faculté au niveau du champ PV. Les simulations effectuées sous l'environnement MATLAB/Simulink ont permis de constituer un dataset (base des données) composé de plus 90 millions d'échantillons dont plus de 30 millions pour chaque classe.

# Chapitre 4

## Détection des défauts

---

### 4.1 Introduction

Dans le chapitre précédent, nous avons collecté les données nécessaires que nous avons prévu utiliser pour la détection d'anomalies. A présent, nous allons présenter et analyser les performances des modèles de classification adoptés (le SVM, le Random Forest et l'arbre de décision) pour les deux types des défauts mentionnés au chapitre précédent.

### 4.2 Pipeline du modèle proposé

Comme nous l'avons présenté au chapitre deux, lors de la conception, nous rappelons ici la méthodologie adoptée qui est basée sur plusieurs étapes successives afin de structurer le processus de traitement et de construction du modèle de classification. Le pipeline du processus est représenté à la Figure 4.1.

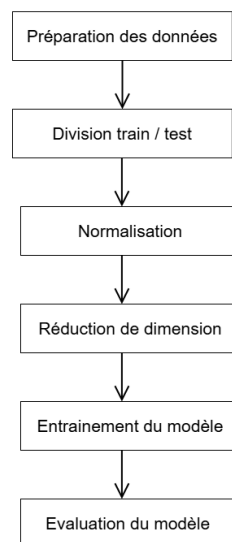


FIGURE 4.1 – Pipeline du modèle

### 4.3 Préparation des données

Les données recueillies dans le chapitre précédent sont rangées dans des fichiers ou chaque état de fonctionnement du système est associée à une étiquette de classe qui indique le type de condition de fonctionnement. Ce rangement des données permet de transformer les résultats de simulation en un problème de classification supervisée, où l'algorithme doit apprendre à reconnaître automatiquement les différentes situations du système.

### 4.4 Division du dataset

Le dataset que nous utilisons dans ce travail provient de la simulation de différents scénarios de fonctionnement du système. Bien que nous ayons initialement généré plus de 3 millions de données par classe pour une simulation de 3 secondes, des contraintes liées aux ressources informatiques limitées nous ont conduits à sélectionner 10 000 échantillons situés dans l'intervalle où nous introduisons le défaut de court-circuit (entre 0,5 s et 0,51 s), correspondant à la zone de chevauchement des trois classes.

Nous avons ensuite divisé le dataset en deux parties : les données d'entraînement (Figure 4.2), représentant 80 % de l'ensemble soit 24 000 données et utilisées pour permettre au modèle d'apprendre les relations entre les variables d'entrée (20 paramètres) et les classes, et les données de test (Figure 4.3), représentant les 20 % restants, utilisées afin d'évaluer sa capacité de généralisation sur des données non vues.

Les statistiques des valeurs reprises sur les Figures 4.2 et 4.3 nous renseignent sur les éléments suivants :

- **Count** : correspond au nombre total d'échantillon du dataset.
- **Mean** : c'est la moyenne des valeurs d'une colonne.
- **Std (standard déviation ou Ecart-type)** : c'est la mesure de dispersion des valeurs autour de la moyenne.
- **Min** : c'est la valeur minimale.
- **Max** : c'est la valeur maximale.
- **25%** : représente la valeur autour de laquelle se situent 25% des données.

- **50%** : c'est la valeur qui sépare les valeurs supérieures et inférieures (médiane).
- **75%** : représente la valeur autour de laquelle se situent 75% des données.

	Time	mVpv	mVpv1	mVpv2	mVpv3	mVpv4	mVpv5	mlpv	mlpv1	mlpv2	...	mlpv4
<b>count</b>	24000.000000	24000.000000	24000.000000	24000.000000	24000.000000	24000.000000	24000.000000	24000.000000	24000.000000	24000.000000	...	24000.000000
<b>mean</b>	0.505005	24.027787	24.027787	23.233875	24.821699	24.027789	24.027785	34.675454	34.675454	34.296752	...	34.675454
<b>std</b>	0.002895	16.924749	16.924749	16.421834	17.510909	16.924747	16.924751	0.456939	0.456939	0.928260	...	0.456939
<b>min</b>	0.500000	0.003933	0.003933	0.003933	0.003933	0.003933	0.003933	34.308445	34.308445	31.992337	...	34.308445
<b>25%</b>	0.502493	0.052979	0.052979	0.052979	0.052979	0.052979	0.052979	34.308455	34.308455	33.898423	...	34.308455
<b>50%</b>	0.505011	35.649333	35.649333	33.010913	36.174730	35.649333	35.649333	34.425098	34.425098	34.308452	...	34.425098
<b>75%</b>	0.507521	36.174719	36.174719	36.174719	37.345700	36.174719	36.174719	35.319744	35.319744	35.319744	...	35.319744
<b>max</b>	0.509999	36.174759	36.174759	36.174759	39.521383	36.174759	36.174759	35.320543	35.320543	35.320543	...	35.320543

FIGURE 4.2 – Extrait statistique données d'entraînement

FIGURE 4.3 – Extrait statistique données test

## 4.5 Normalisation des données

Avant d'entraîner le modèle, nous avons mis les données sur la même échelle afin d'éviter que les données avec les plus grandes valeurs ne dominent le processus d'apprentissage et ainsi fausser le résultat du modèle. Pour y parvenir, la méthode de standardisation a été utilisée et elle consiste à transformer chaque variable de manière qu'elle possède une moyenne égale à 0 et un écart-type égal à 1.

La transformation se fait suivant l'équation ci-dessous :

$$X_{norm} = \frac{X - \mu}{\sigma} \quad (4.1 \text{ [[45]])}$$

Avec  $X$  la valeur de la variable,  $\mu$  la moyenne de la variable dans le dataset,  $\sigma$  l'écart-type de la variable,  $X_{norm}$  la valeur normalisée.

## 4.6 Réduction de dimension par PCA

Dans le but de réduire la complexité du dataset et améliorer l'efficacité du modèle, nous avons utilisé la technique de réduction de dimension appelée Analyse en Composantes Principales (PCA).

L'Analyse en composantes principales (PCA) a permis de transformer les variables initiales en un nouvel ensemble de variables, appelées composantes principales. Ces dernières conservent l'essentiel de l'information contenue dans les données d'origine, tout en réduisant significativement la dimensionnalité du jeu de données.

Cette réduction présente plusieurs avantages, notamment la diminution du temps de calcul ainsi que la réduction du bruit présent dans les données.

## 4.7 Entraînement du modèle

Nous entraînons nos modèles sur un ensemble de 24 000 échantillons pour toutes les classes (Figure 4.2) et pour chaque modèle nous prenons en compte certains paramètres qui nous permettent d'améliorer les performances des modèles :

- Pour le **SVM**, nous utilisons le noyau de type *Radial Basis Function* (RBF) qui est adapté aux relations non linéaires [46] ; nous utilisons également les paramètres  $\gamma$  et  $c$  respectivement *scale* et 1.
- Quant au **Decision Tree**, nous utilisons les paramètres par défaut permettant de construire automatiquement une structure de décision adaptée aux données bien que cela accentue la sensibilité au surapprentissage.
- Enfin, le **Random Forest** utilise le nombre d'arbres de décision  $n = 100$ , où chaque arbre est entraîné sur une partie des données et le résultat final réside d'un vote majoritaire.

## 4.8 Présentation des résultats

Après l'apprentissage, le modèle a été évalué sur les données de test afin de mesurer ses performances à l'aide de la matrice de confusion et des métriques (*Accuracy*, Précision, Rappel et F1-score).

Nous avons testé les modèles sur 6000 données en raison de 2000 données par classe. Les résultats se présentent comme suit :

**Modèle SVM** La Figure 4.4 nous montre clairement que :

- Le modèle a classé correctement les 2000 échantillons de classe normale (0) sans les confondre avec les autres classes : le modèle identifie le fonctionnement normal avec une précision élevée.
- Pour la classe du défaut de court-circuit (1) : sur 2000 cas, 1999 cas ont été correctement prédits et seul un cas prédit comme étant normal : la détection est presque bien faite avec une très faible erreur.
- Parmi les 2000 échantillons du défaut d'ombrage partiel (2), 1898 échantillons sont correctement classés et 102 sont confondus avec la classe normale. Ce qui précède renseigne que le modèle a plus de difficulté à distinguer certaines situations.

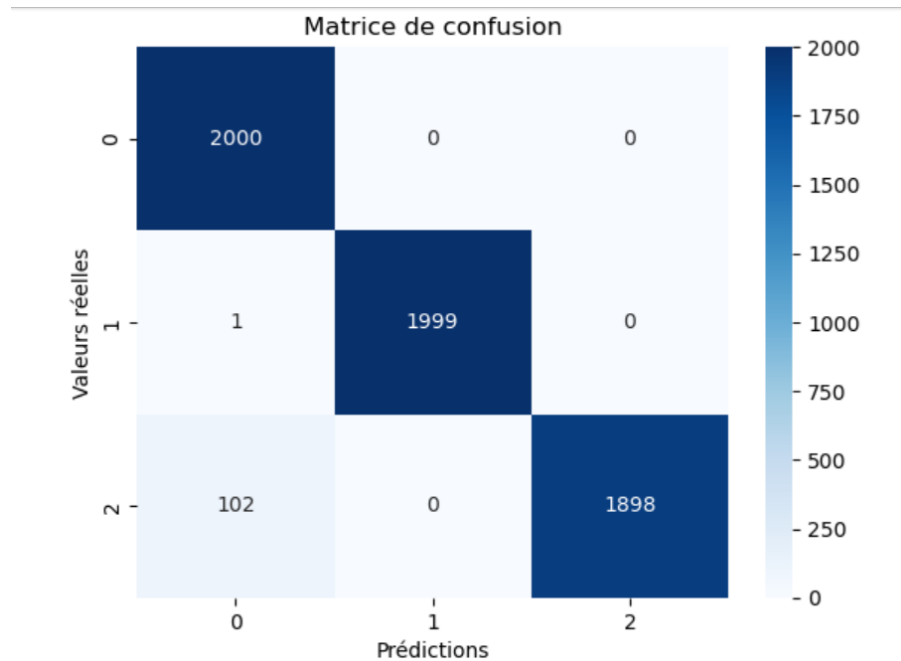


FIGURE 4.4 – Matrice de confusion du SVM

L'analyse du rapport de classification du SVM à la Figure 4.5 nous révèle une performance du modèle avec une exactitude (*accuracy*) de 98,28 % sur un ensemble de 6000 des données test, soit environ 5897 instances correctement classées. Le court-circuit est parfaitement détecté : pratiquement toutes les instances de cette classe ont été reconnues. Par contre, pour l'ombrage partiel, environ 51 cas sur 1000 n'ont pas été correctement détectés : le modèle a une difficulté à distinguer cette classe.

	precision	recall	f1-score	support
0	0.951022	1.000000	0.974896	2000
1	1.000000	0.999500	0.999750	2000
2	1.000000	0.949000	0.973833	2000
accuracy			0.982833	6000
macro avg	0.983674	0.982833	0.982826	6000
weighted avg	0.983674	0.982833	0.982826	6000

FIGURE 4.5 – Rapport de classification du SVM

**Modèle Decision Tree** La matrice de confusion de la Figure 4.6 montre que :

- Le modèle a été capable de classer 1997 sur 2000 cas comme normal et il a confondu 3 cas normaux à l’ombrage partiel.
- Sur 2000 cas de court-circuit, le modèle en a classé correctement 1999 et un seul est confondu à l’ombrage partiel.
- Pour l’ombrage partiel, 1994 cas sont classés correctement et 6 sont confondus à la classe normale.

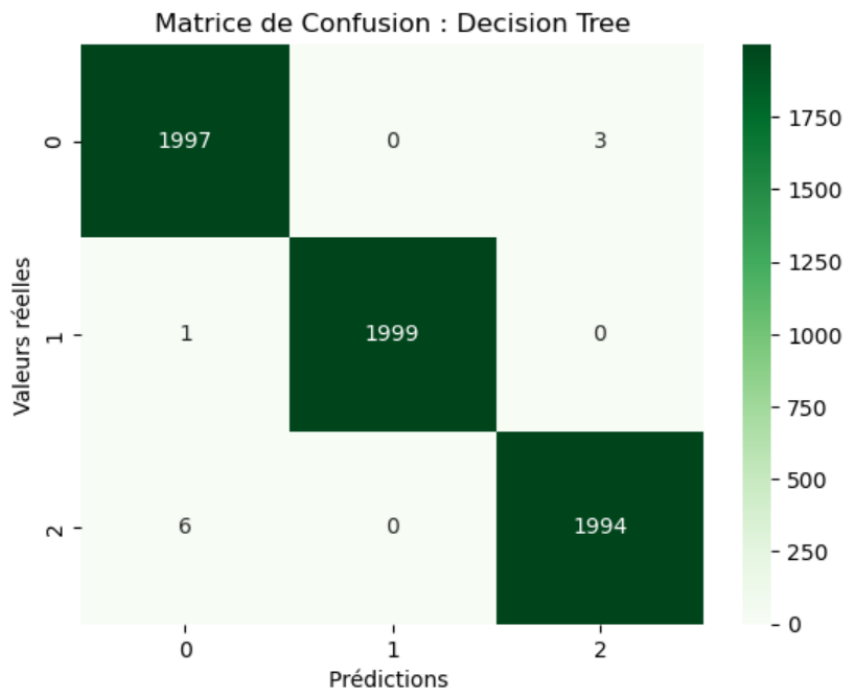


FIGURE 4.6 – Matrice de confusion de l’arbre de décision

Les scores proches de 1 nous renseignent que le modèle peut distinguer les différentes

classes sans trop d'erreurs. À partir de ce que nous avons vu à la matrice de confusion qui montre seulement 10 échantillons sur 6000 qui ont été mal classés, ce résultat conduit aux valeurs que nous lisons à la Figure 4.7.

	precision	recall	f1-score	support
0	0.996507	0.998500	0.997502	2000
1	1.000000	0.999500	0.999750	2000
2	0.998498	0.997000	0.997748	2000
accuracy			0.998333	6000
macro avg	0.998335	0.998333	0.998334	6000
weighted avg	0.998335	0.998333	0.998334	6000

FIGURE 4.7 – Rapport de classification du Decision Tree

**Modèle Random Forest** La Figure 4.8 nous donne les informations suivantes :

- Le modèle a correctement classé 1998 cas normaux (0) et a fait une confusion de 2 cas normaux avec l'ombrage partiel (2).
- Sur un ensemble de 2000 cas de court-circuit (1), 1999 ont été correctement classés et un seul confondu à l'ombrage partiel (2).
- Parmi les 2000 échantillons du défaut d'ombrage partiel (2), 1995 échantillons sont correctement classés et 5 sont confondus avec la classe normale (0).

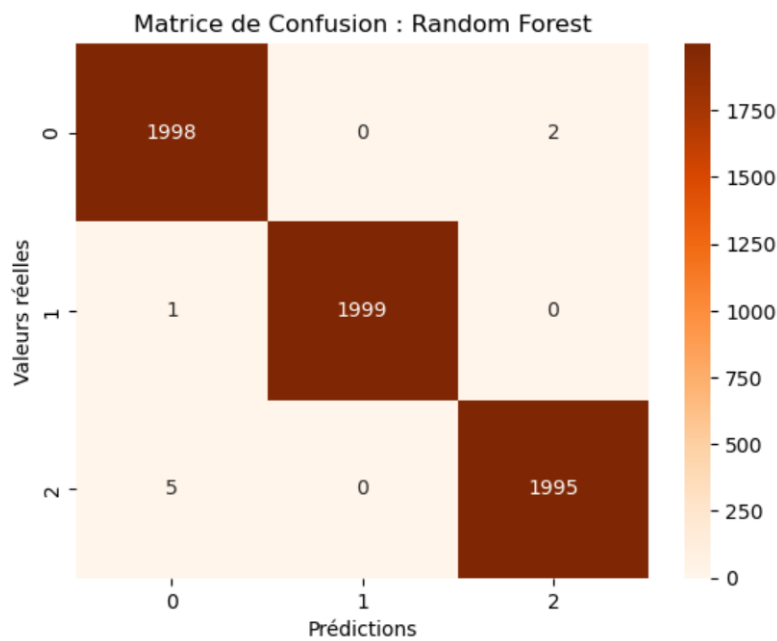


FIGURE 4.8 – Matrice de confusion de Random Forest

Sur l'ensemble des données ce modèle peut être capable de détecter les défauts sans trop d'erreur comme illustré sur la Figure 4.9.

	precision	recall	f1-score	support
0	0.997006	0.999000	0.998002	2000
1	1.000000	0.999500	0.999750	2000
2	0.998998	0.997500	0.998249	2000
accuracy			0.998667	6000
macro avg	0.998668	0.998667	0.998667	6000
weighted avg	0.998668	0.998667	0.998667	6000

FIGURE 4.9 – Rapport de classification du Random Forest

## 4.9 Discussion sur les résultats

Les résultats de la section nous permettent de voir le comportement de chaque modèle par rapport à chaque classe :

- Les trois modèles sont parvenus à mieux identifier 1999 cas avec le même résultat pour la classe court-circuit, ce qui est bon pour la sécurité de l'installation.
- Le SVM a une difficulté à classer les défauts d'ombrage partiel avec 102 cas classés

normaux.

- Le Random Forest et le Decision Tree présentent les résultats équilibrés et presque identiques. Cependant le Random Forest prédit avec moins de confusion sur l'ensemble de classes.

<b>Modèle</b>	<b>Accuracy</b>	<b>Précision</b>	<b>F1-Score</b>
<b>SVM</b>	<b>0.982833</b>	<b>0.983674</b>	<b>0.982826</b>
<b>Random Forest</b>	<b>0.998667</b>	<b>0.998668</b>	<b>0.998667</b>
<b>Decision Tree</b>	<b>0.998333</b>	<b>0.998335</b>	<b>0.998334</b>

FIGURE 4.10 – Performances globales des modèles

La Figure 4.10 présente une synthèse des performances des modèles. Sur les 6000 instances du jeu de test, le Random Forest en a correctement classé 5988, l'arbre de décision 5989 et le SVM 5897. Ces résultats confirment que le Random Forest est le modèle le plus performant suivi de l'arbre de décision et après vient le SVM.

## 4.10 Conclusion partielle

Dans ce chapitre, nous avons détaillé l'entraînement ainsi que l'évaluation des trois modèles de Machine Learning notamment le SVM, le Decision Tree et le Random Forest.

Les résultats obtenus à partir des matrices de confusion et des différentes métriques de performance ont permis d'évaluer de manière détaillée chaque modèle. L'analyse comparative met en évidence que le Random Forest présente les meilleures performances globales. Le Support Vector Machine (SVM), bien que performant, éprouve des difficultés dans la détection de l'ombrage partiel, tandis que le Decision Tree affiche des performances légèrement inférieures.

Toutefois, ces modèles nécessitent d'être validés sur des données issues de systèmes photovoltaïques réels, non simulés, afin de mieux refléter les conditions d'exploitation, généralement caractérisées par la présence de bruit. Par ailleurs, l'étude s'est limitée à deux types de défauts, ce qui pourrait restreindre la capacité de généralisation des modèles développés.

# Conclusion générale

---

Ce travail avait pour objectif de proposer une approche technologique capable d'assurer une détection intelligente et précoce des anomalies dans les systèmes photovoltaïques à partir des données de fonctionnement. La problématique principale consistait à identifier une méthode fiable permettant de surveiller automatiquement une installation PV et de détecter efficacement les défauts susceptibles d'affecter ses performances.

Les résultats obtenus montrent que l'utilisation de modèles d'apprentissage supervisé constitue une solution pertinente pour la classification des états du système (fonctionnement normal, ombrage partiel et court-circuit). Le modèle basé sur les Machines à Vecteurs de Support (SVM) a été étudié et d'autres approches telles que Random Forest et Décision Tree ont également été évaluées, leurs matrices de confusion ainsi que leurs métriques ont été présentés; une analyse comparative met en évidence la capacité des différentes méthodes à assurer une détection fiable des anomalies étudiées.

La modélisation du système sous MATLAB/Simulink, la génération d'un dataset annoté, ainsi que l'entraînement et l'évaluation des modèles ont permis de démontrer la faisabilité d'une solution de surveillance intelligente appliquée aux systèmes photovoltaïques. Ce travail contribue ainsi à l'amélioration des stratégies de maintenance prédictive et à la réduction des pertes de production énergétique.

Cependant, malgré ces résultats prometteurs, certaines limites doivent être soulignées. L'étude repose principalement sur des données simulées, qui ne reproduisent pas entièrement la complexité d'un environnement réel, notamment en ce qui concerne le vieillissement des équipements, les variations climatiques imprévisibles et le bruit de mesure. Par ailleurs, seuls deux types de défauts ont été considérés et le système est étudié à circuit ouvert, ce qui limite la généralisation du modèle à l'ensemble des anomalies susceptibles d'affecter une installation photovoltaïque. Une analyse comparative plus approfondie, intégrant d'autres architectures d'apprentissage et des tests en conditions réelles, permettrait de consolider davantage les conclusions obtenues.

Dans la perspective de travaux futurs, il serait pertinent de valider les modèles sur des données issues d'une centrale PV réelle afin d'évaluer leur robustesse en conditions opérationnelles. L'intégration d'autres types de défauts, tels que la dégradation progressive des

modules ou les défauts de câblage, permettrait également d'améliorer la généralisation du système. De plus, l'exploration de méthodes d'apprentissage plus avancées pourrait optimiser davantage les performances de classification.

# Bibliographie

---

- [1] B. LONG, *Detection et Localisation des défauts dans un système PV*. France, 2011.
- [2] A. L., G. S. et R. A. NELSON, *Handbook of Photovoltaic Science and Engineering*. John Wiley & Sons, Ltd, 2011.
- [3] J.-P. KLEIDER, *Effet photovoltaïque : principes physiques*, <https://books.openedition.org/editions-cnrs/person/10924>, 2011.
- [4] B. K. AHMED BEHALIL MOHCÈNE, “Etude et comparaison des différentes configurations des panneaux photovoltaïques sous l’effet de l’ombrage partiel”, mém.de mast., 2019.
- [5] *Petite cellule solaire 1 v 200 ma sol3n814241*, <https://fundatia-adept.org/PETITE-CELLULE-SOLAIRE-1-V-200-MA-SOL3N814241>.
- [6] H. KAMELIA, “Modélisation d’une cellule photovoltaïque”, mém.de mast., 2012.
- [7] S. MOTAHHIR, A. E. GHZIZAL et A. DEROUICH, “Modélisation et commande d’un panneau photovoltaïque dans l’environnement psim”, in *Congrès International de Génie Industriel et Management des Systèmes*, hal-01351493, Maroc.
- [8] G. B. SAHOUR DOUNIA, “Etude et modélisation d’un panneau photovoltaïque”, mém.de mast., Algérie, 2021.
- [9] B. HAMZA, “Modelisation et commande d’un panneau”, Memoire de master, Département d’électronique, filiere d’automatique.
- [10] UNIVERSITE IBN KHALDOUN DE TIARET, *Modelisation de la cellule photovoltaïque*, Algérie.
- [11] *Civisol.fr*, [En ligne], 2012.

- 
- [12] H. KAMELIA, “Modelisation d’une cellule photovoltaïque : etude comparative”, Memoire de Master en electrotechnique, Option Machines electriques, Algerie.
- [13] E. H. SEPÚLVEDA-OVIEDO, “Detection and diagnosis of faults and performance losses in high-power photovoltaic power plants”, 2025.
- [14] E. H. S. OVIEDO, “Detection and diagnosis of faults and performance losses in highpower photovoltaic power plants”, thèse de doct., École doctorale et spécialité : Génie Electrique, Toulouse, 2023.
- [15] A. B., “Identification et detection des defauts dans les installations photovoltaïque”, thèse de doct., Université de Pepignan Via Domitia, France, juin 2021.
- [16] O. ALOULOU, “Utilisation des réseaux des neurones artificiels pour l’optimisation de la commande mppt et la détection des défauts des panneaux solaires”, mém.de mast., Université du Québec, 2024.
- [17] B. N., *Université badji mokhtar-annaba*, 2024.
- [18] B. M. SAMIR, “Etude d’un convertisseur boost pour le systeme photovoltaïque”, mém.de mast., Spécialité : Energie et environnement, 2020.
- [19] MESBAHI, 2015.
- [20] *Voltania*, [En ligne], oct. 2006.
- [21] B. K. BOUZANA SANAA, “Diagnostic de pannes dans un système photovoltaïque”, mém.de mast., 2022.
- [22] *Researchgate.fr*, 2013.
- [23] *Researchgate.fr*, [En ligne ; Accès le 21 novembre 2025], 2013.
- [24] M. Z. SOUAD, “Réalisation d’un régulateur solaire à base de microcontrôleur pour le contrôle de l’état de charge et la protection des accumulateurs”, mém.de mast., 2008.

- [25] A. M. ALI, C. J. et F. H., “Diagnostic des systèmes dynamiques linéaires sans modèle explicite”, *Journal Européen des Systèmes Automatisés (JESA)*, t. 45, p. 339–362, jan. 2011.
- [26] O. GUENOUNOU, “Méthodologie de conception de contrôleurs intelligents par l’approche génétique. application à un bioprocédé”, thèse de doct., Université Paul Sabatier, Toulouse, 2009.
- [27] N. AOUCHICHE, “Defauts lies aux systemes photovoltaïques autonomes et techniques de diagnostic”, *Revue des énergies Renouvelables*, t. 21, n° 2, p. 247–265, 2018.
- [28] *Schema global de detection et de diagnostic des defauts*, [https://www.researchgate.net/figure/Schema-global-de-detection-et-de-diagnostic-des-defauts\\_fig4\\_328577571](https://www.researchgate.net/figure/Schema-global-de-detection-et-de-diagnostic-des-defauts_fig4_328577571), 2018.
- [29] B. SAMIA, *Surveillance et diagnostic de défauts dans les systèmes électriques*, 2018.
- [30] I. SASSI, “Diagnostic de défauts dans un panneau photovoltaïque étude théorique et pratique”, mém.de mast., Tunis, 2017.
- [31] S. T. KEBIR, “A comprehensive study of diagnosis faults”, *Engineering Review*, 2020.
- [32] G. JAMES, D. WITTEN, T. HASTIE et R. TIBSHIRANI, *An introduction to statistical learning*. 2017.
- [33] R. S. SUTTON et A. G. BARTO, *Reinforcement learning : An introduction*. 2018.
- [34] N. B. MARIVAUX, “Contribution a la detection des defauts electriques dans les moteurs electriques inductifs en utilisant les techniques avancees de machine learning”, mém.de mast., Kinshasa, 2023.

- [35] B. S. TCHAMBA KUINZE, “Comparaison de l’efficacité du deep learning et de l’apprentissage automatique classique pour la reconnaissance d’activités humaine dans les habitats intelligents”, Mémoire de maîtrise (M. Sc.) Université du Québec à Chicoutimi, Québec, Canada, 2024.
- [36] *Matrice de confusion*, <https://datascientest.com/>.
- [37] *Matrice de confusion*, <https://liora.io/matrice-de-confusion>.
- [38] K. N. S. MUTAONA, “Detection and classification of shading and short-circuit faults in photovoltaïque systeme using svm, dt and rf with principal component analysis”,
- [39] N. M. SAMUEL, *Detection et localisation des defauts dans un systeme pv par extreme machine learning*, 2025.
- [40] A. F. AMIRI, H. O., A. C. et S. K., “Design and analysis of pv systems based on machine learning using random forest classifier”, *E. C. and M.*, t. 3, 2023.
- [41] C. CHEN, P. J., L. J., Z. Z., C. C., Y. H., L. L., J. W., S. Y. et C. L., “Fault diagnosis and classification for photovoltaic arrays based on principal component analysis and support vector machine”, in *IOP Conference Series : Earth and Environmental Science*, t. 188, 2018, p. 012 089.
- [42] T. K. ROY, A. M. T. OO et S. K. GHOSH, *Designig high-order sliding mode controller for photovoltaic and batterie energy storage system-based dc micro-grids with ann-mppt*, 2024.
- [43] R. M. et A. RAZA, “A support vector machine learning-based protection technique for mt-hvdc systems”, *Energies*, 2020.
- [44] H. FU et K. QI, “Evaluation model of teachers’ teaching ability based on improved random forest with grey relation projection”, *Hindawi Scientific Programming*, 2022.

- 
- [45] A. S. SRINIVAS, Y. S., V. K., I. D. et S. T., “Data standardization : key to effective data integration”, *Advanced Innovations in Computer Programming Languages*,
- [46] Q. L. HU, C. et Z., “Feature selection for support vector machines with rbf kernel”, *Springer Science+Business Media B.V.*, 2011.
- [47] P. J. M. ALI, R. H. F., D. N. et S. A., “Machine learning technical reports”, rapp. tech., 2014.

# Annexes

---

## Annexe A : Fonction Matlab de collecte des données

```
function logDataToCSV(mVpv, mVpv1, mVpv2, mVpv3, mVpv4, mVpv5,...
                    mIpv, mIpv1, mIpv2, mIpv3, mIpv4, mIpv5,...
                    Ist1, Ist2, Ist3, Itot, Vtot, Ptot,...
                    I_bat, V_bat)

%#codegen

persistent fileID counter;
sampleTime = 0.000001;

if isempty(fileID)
    filename = 'D:\MES DOCUMENTS\Mes_Docs\Tfcaan\Travail_en_Cours_Ann-Joa\Simulink\
modélisation\SimulationPVTFE\simulation_anntfe22_data.csv';
    fileID = fopen(filename, 'w');

    % EN-TÊTE : On utilise des espaces ou une largeur fixe pour que
    % les titres soient alignés avec les données en dessous.
    fprintf(fileID, 'Time      ;mVpv      ;mVpv1      ;mVpv2      ;mVpv3      ;mVpv4
;mVpv5      ;mIpv      ;mIpv1      ;mIpv2      ;mIpv3      ;mIpv4      ;mIpv5      ;Ist1
;Ist2      ;Ist3      ;Itot      ;Vtot      ;Ptot      ;Ibat      ;Vbat\n');
    counter = 0;
end

currentTime = counter * sampleTime;

% DONNÉES : %10.6f garantit que chaque nombre occupe exactement 10 caractères.
% Cela force l'alignement vertical peu importe la valeur du chiffre.
formatSpec = '%10.8f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f
f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f;%10.6f\n';

fprintf(fileID, formatSpec, ...
        currentTime, ...
        mVpv, mVpv1, mVpv2, mVpv3, mVpv4, mVpv5, ...
        mIpv, mIpv1, mIpv2, mIpv3, mIpv4, mIpv5, ...
        Ist1, Ist2, Ist3, Itot, Vtot, Ptot, ...
        I_bat, V_bat);

counter = counter + 1;
end
```

FIGURE 11 – Capture d'écran du script Matlab pour la collecte des données

## Annexe B : Extrait du code de Machine Learning

```
df_normal = load_specific_data('FonctionnementNormal', 0) # Label 0: Normal
df_cc = load_specific_data('Court-circuit', 1) # Label 1: Court-circuit
df_ombrage = load_specific_data('Ombrage partielle', 2) # Label 2: Ombrage

# Fusion et mélange aléatoire pour éviter les biais
data = pd.concat([df_normal, df_cc, df_ombrage]).sample(frac=1, random_state=42)
data = data.dropna() # nettoyage des éventuelles valeurs manquantes

# Préparation des données
X = data.drop('label', axis=1)
y = data['label']

# Division 80% Entraînement / 20% Test
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y)

# Normalisation
# Standardise les variables (Tension, Courant, etc.) à une moyenne de 0 et variance 1
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# Réduction de dimensionnalité par PCA
# réduction des 21 dimensions d'origine à 5 composantes principales
# Cela permet de conserver 95% de la variance tout en réduisant le temps de calcul par 2
pca = PCA(n_components=5)
X_train_pca = pca.fit_transform(X_train_scaled)
X_test_pca = pca.transform(X_test_scaled)
# hyperparamètres : C = 1.0 et gamma
svm_model = SVC(kernel='rbf', C=1.0, gamma='scale', random_state=42)
svm_model.fit(X_train_pca, y_train)

# 6. Evaluation des performances
y_pred = svm_model.predict(X_test_pca)
numeric_labels = [0, 1, 2]
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=numeric_labels,
            yticklabels=numeric_labels)

plt.xlabel('Prédictions')
plt.ylabel('Vraies valeurs')
plt.title('Matrice de confusion')
plt.show()

print("--- Rapport de Classification (SVM) SVM ANNOZI ---")
print(classification_report(y_test, y_pred,
                           target_names=['0', '1', '2'],
                           digits=3)) # Augmentez ce chiffre selon votre besoin

except FileNotFoundError as e:
    print("Erreur : (4)")
```

FIGURE 12 – Extrait de l'implémentation du modèle de Machine Learning sous Python

## Annexe C : figure du système étudié

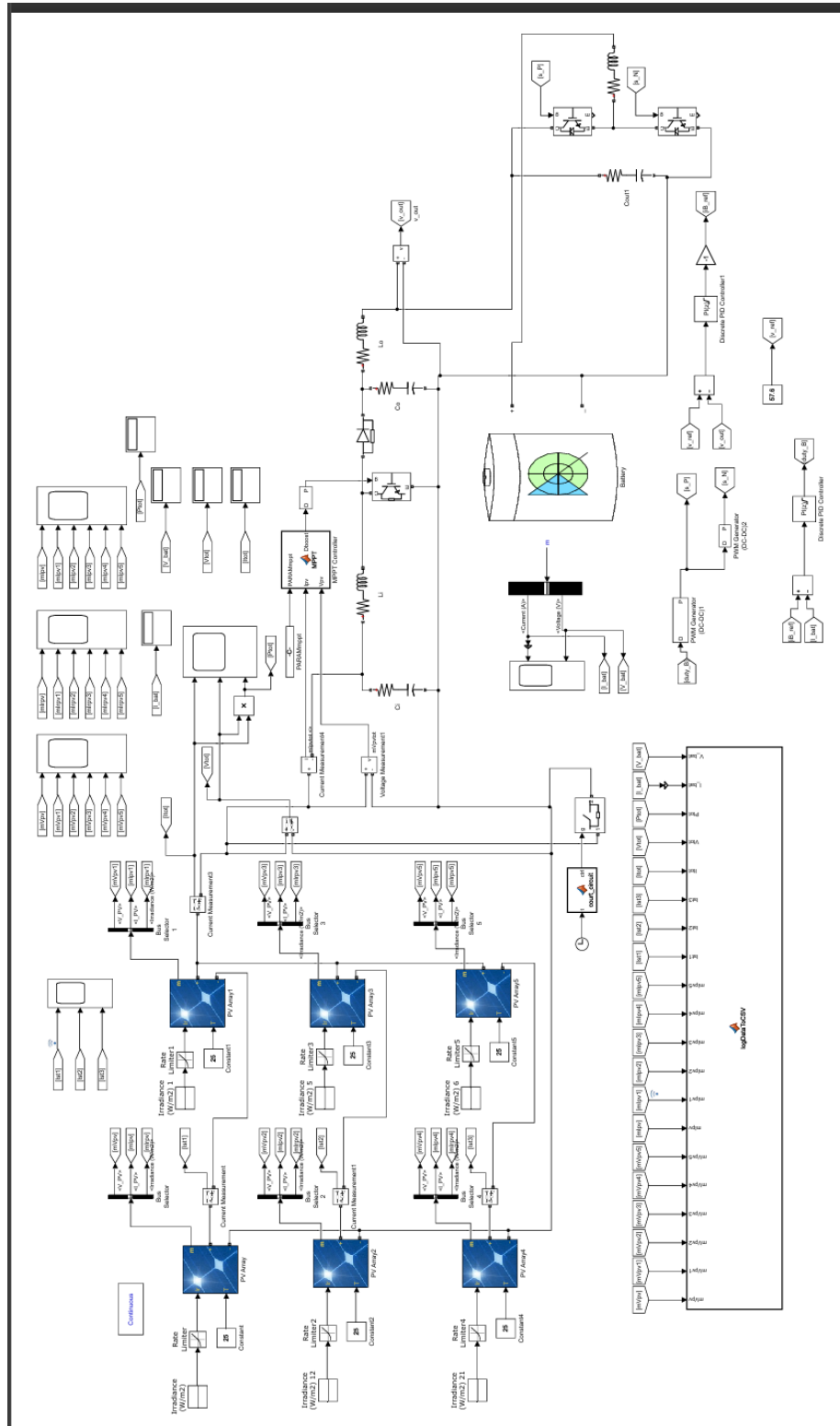


FIGURE 13 – Modèle Simulink du système choisi