



People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Echahid Hamma Lakhdar University - El Oued



Faculty of Technology
Department of Electrical Engineering

Domain: **Science and Technology**
Division: **Electrical Engineering**
Specialty: **Electrical Networks**

ACADEMIC MASTER

Presented by:

- 1. KERBATI Maroua.**
- 2. CHAUCHE Roufaida.**

Data-Driven Statistical Methods for Fault Detection in Solar Photovoltaic Systems

Dissertation Submitted in Partial Fulfillment of the Requirements for the Master

Degree in Electrical Networks

Publicly defended in: 13/06 /2026

Board of Examiners:

Dr. Idriss BABA ARBI	President	Univ Eloued
Dr. ILyes CHAOUKI	Examiner	Univ Eloued
Dr. Riadh TOUMI	Supervisor	Univ Eloued

Academic Year: 2025/2026

الشكر والعرفان

الحمد لله رب العالمين،

والصلاة والسلام على أشرف المرسلين وسيد الخلق أجمعين،

وعلى آله وصحبه ومن تبعهم بإحسان إلى يوم الدين.

أما بعد، نتقدم نحن الطالبتان بخالص الشكر والتقدير والامتنان إلى أستاذنا المشرف الدكتور رياض تومي، الذي حظينا بفرصة العمل تحت إشرافه، ولم يخل علينا بذائعه السديدة وتوجيهاته العلمية القيمة طيلة فترة إنجاز هذا البحث. ونحن نشمن عالياً رحابة صدره، ومتابعته المستمرة لنا، وحرصه على توجيهنا خطوة بخطوة في مرحلتي الدراسة والمحاكاة لتقديم هذا العمل في أفضل صورة ممكنة.

نسأل الله العليّ القدير أن يبارك في علمه وعمله، وأن يوفقه دوماً في مسيرته الأكاديمية والمهنية، ويجعل هذا الجهد في ميزان الحسنات



الاهداء



بسم الله الرحمن الرحيم

الحمد لله الذي وفقني وأعانني على إتمام هذا العمل المتواضع.

أهدي ثمرة جهدي هذا إلى أغلى الناس على قلبي:

إلى أمي الحبيبة...

مصدر الحنان والعطاء، وسرّ القوة والثبات، التي كانت دعواتها الصادقة ترافقني في كل خطوة،

فجزاها الله عني خير الجزاء وأدامها نعمة لا تزول.

إلى أبي العزيز...

من كان لي السند والعون، ومن علمني أن الاجتهاد والصبر طريق النجاح، فكان مثلاً للعطاء

والتضحية، حفظه الله وأطال في عمره.

إلى إخوتي وأخواتي...

الذين شاركوني لحظات التعب والفرح، وكانوا خير داعم ومشجع طوال مسيرتي الدراسية.

إلى عائلتي الكريمة وكل من وقف معي إلى جانبي...

وساندني بكلمة طيبة أو دعوة صادقة أو نصيحة نافعة.

وإلى كل أساتذتي الأفاضل...

الذين أناروا لي طريق العلم والمعرفة، وأسهموا في تكويني العلمي.

أسأل الله عز وجل أن يجعل هذا العمل خالداً لوجهه الكريم، وأن يوفقنا جميعاً لما فيه الخير والنجاح.

والحمد لله رب العالمين.

مروة



الاهداء

"من قال أنا لها نالها، وأنا لها وإن أبى رخصاً عنها أتيت بها"

لم تكن الرحلة قصيرة، ولم يكن الطريق ممهداً بالسمولة التي تمنيتها، لكن بالإرادة والصبر والعزيمة وصلت
إلى هذه اللحظة.

إلى الذي زين اسمي بأجمل الألقاب، ومن دعمني دون حدود، وكان سدي وقوتي في كل مراحل حياتي،
...وغرس في نفسي حب العلم والاجتهاد

إلى أبي الغالي.

إلى من جعل الله الجنة تحت قدميها، وإلى القلب الذي احتواني حباً وحناناً، وإلى من كانت دعواتها
...تدافقني في كل خطوة، وكانت نوراً يضيء طريقي في أصعب الأوقات

إلى أمي العزيزة

إلى إخوتي وأخواتي الذين كانوا خير سند وعمود إلى جميع أفراد عائلتي الكريمة، الذين أحاطوني بالمحبة
والتشجيع والدعاء الصادق

إلى صديقتي مروة، رفيقة سنوات جميلة وشريرة في هذا العمل شكراً لرفقتك الجميلة ودعمك الدائم
أتمنى لك دوام النجاح والتوفيق.

أهديكم هذا الإنجاز وثمرة سنوات من الجهد والمثابرة، راجية من الله أن يجعل فيه الخير والتوفيق، وأن
يوفقني لما هو أفضل في القادم من الأيام

فالحمد لله على ما مضى، والحمد لله على ما تحقق، والحمد لله على ما هو آتٍ.

رفيدة

الملخص:

يسعى هذا المشروع، من خلال إنشاء إطار عمل متطور للمراقبة والتشخيص الإحصائي القائم على البيانات، إلى تحسين فعالية وموثوقية الأنظمة الكهروضوئية (PV) العاملة في الظروف الصحراوية. الهدف الرئيسي من الدراسة هو إجراء مقارنة كمية ونوعية شاملة بين طريقة تحليل المكونات الرئيسية الساكنة التقليدية (PCA) وامتدادها الديناميكي، تحليل المكونات الرئيسية الديناميكي (DPCA)، والذي تم تنفيذه في بيئة ماتلاب (MATLAB) للكشف المبكر عن الأعطال وعزلها. تم تقييم المنهجية المقترحة باستخدام بيانات تشغيلية وجوية حقيقية تم جمعها من محطة الطاقة الشمسية "وادي الكبريت" في ولاية سوق اهراس، الجزائر، وذلك من أجل إعطاء تطبيق عملي. وأظهرت نتائج المحاكاة أن أسلوب (DPCA) متطور تقنياً بشكل واضح مقارنة بطريقة (PCA) الساكنة التقليدية. ومن خلال استخدام حدود التحكم الإحصائي القائمة على مؤشرات "هوتيلينج T^2 " ومربع خطأ التنبؤ (SPE)، خفضت طريقة (DPCA) بشكل كبير من معدل الإنذارات الكاذبة (FAR) بينما زادت من معدل الكشف عن الأعطال (FDR) عن طريق دمج الديناميكيات الزمنية في إطار المراقبة. بالإضافة إلى ذلك، أظهرت مخططات المساهمة القائمة على إعادة البناء (RBC) درجة عالية من الفعالية في تحديد العوامل المسببة للسلوك غير الطبيعي للنظام بدقة، مما يسهل خطط الصيانة الاستباقية ويقلل من خسائر الطاقة. وتؤكد النتائج إمكانات أساليب المراقبة الإحصائية الديناميكية متعددة المتغيرات كأدوات موثوقة لتعزيز إدارة الأعطال والأداء التشغيلي للأنظمة الشمسية التي تعمل في البيئات الصحراوية القاسية.

الكلمات المفتاحية: الأنظمة الكهروضوئية (PV)، تشخيص الأعطال، تحليل المكونات الرئيسية (PCA)، تحليل المكونات الرئيسية الديناميكي (DPCA)، محطة وادي الكبريت للطاقة الكهروضوئية، حدود التحكم الإحصائي، الكشف عن الأعطال وعزلها. (FDI).

Abstract:

By creating a sophisticated data-driven statistical monitoring and diagnostic framework, this project seeks to improve the effectiveness and dependability of photovoltaic (PV) systems operating in desert conditions. The main goal of the study is to perform a thorough quantitative and qualitative comparison between the traditional static Principal Component Analysis (PCA) method and its dynamic extension, Dynamic Principal Component Analysis (DPCA), which is implemented in the MATLAB environment for early fault detection and isolation.

The suggested methodology was assessed using actual operational and meteorological data gathered from the Oued El Kebrit solar power plant in Souk Ahras Province, Algeria, in order to give practical applicability. The simulation results showed that the DPCA approach is clearly more technically advanced than the traditional static PCA method. By using statistical control limits based on Hotelling's T^2 and Squared Prediction Error (SPE) indices, DPCA considerably decreased the False Alarm Rate (FAR) while increasing the Fault Detection Rate (FDR) by integrating temporal dynamics into the monitoring framework.

Additionally, Reconstruction-Based Contribution (RBC) charts demonstrated a high degree of efficacy in precisely pinpointing the factors causing aberrant system behaviour, facilitating proactive maintenance plans and reducing energy losses. The results validate the potential of dynamic multivariate statistical monitoring approaches as dependable instruments for enhancing the fault management and operational performance of solar systems operating in harsh desert environments.

Keywords: Photovoltaic Systems (PV), Fault Diagnosis, Principal Component Analysis (PCA), Dynamic Principal Component Analysis (DPCA), Oued El Kebrit Photovoltaic Power Plant, Statistical Control Limits, Fault Detection and Isolation (FDI).

Résumé:

En créant un cadre sophistiqué de surveillance et de diagnostic statistique basé sur les données, ce projet vise à améliorer l'efficacité et la fiabilité des systèmes photovoltaïques (PV) fonctionnant dans des conditions désertiques. Le but principal de l'étude est de réaliser une comparaison quantitative et qualitative approfondie entre la méthode traditionnelle de l'Analyse en Composantes Principales (PCA) statique et son extension dynamique, l'Analyse en Composantes Principales Dynamique (DPCA), implémentée dans l'environnement MATLAB pour la détection précoce et l'isolation des défauts.

La méthodologie suggérée a été évaluée à l'aide de données opérationnelles et météorologiques réelles collectées auprès de la centrale solaire d'Oued El Kebrit dans la wilaya de Souk Ahras, en Algérie, afin d'offrir une applicabilité pratique. Les résultats de la simulation ont montré que l'approche DPCA est nettement plus avancée sur le plan technique que la méthode PCA statique traditionnelle. En utilisant des limites de contrôle statistique basées sur les indices T^2 de Hotelling et de l'Erreur Quadratique de Prédiction (SPE), la DPCA a considérablement réduit le Taux de Fausses Alarmes (FAR) tout en augmentant le Taux de Détection des Défauts (FDR) par l'intégration de la dynamique temporelle dans le cadre de surveillance.

De plus, les graphiques de Contribution Basée sur la Reconstruction (RBC) des défauts ont démontré un haut degré d'efficacité pour cibler précisément les facteurs causant le comportement anormal du système, facilitant ainsi les stratégies de maintenance proactive et réduisant les pertes d'énergie. Les résultats valident le potentiel des approches de surveillance statistique multivariée dynamique en tant qu'outils fiables pour améliorer la gestion des défauts et les performances opérationnelles des systèmes solaires fonctionnant dans des environnements désertiques rigoureux.

Mots-clés : Systèmes Photovoltaïques (PV), Diagnostic des Défauts, Analyse en Composantes Principales (PCA), Analyse en Composantes Principales Dynamique (DPCA), Centrale Photovoltaïque d'Oued El Kebrit, Limites de Contrôle Statistique, Détection et Isolation des Défauts (FDI).

Table of Contents

الشكر والعرفان.....	1
الاهداء.....	2
الاهداء.....	3
الملخص:	4
Abstract:	5
Résumé:	6
Table of Contents	7
List of Tables	10
List of figures.....	11
List of Abbreviations	12
List of Notations & Symbols	12
General Introduction.....	1

CHAPTER I

Fundamentals of Solar Photovoltaic Systems

1.1 Introduction.....	3
1.2 Overview of Solar Photovoltaic Systems	3
1.2.1 Types of Solar Power Systems:	3
1.2.2 Advantages and Challenges of PV Systems	5
1.3 Photovoltaic Fundamentals.....	7
1.3.1 PV Cell Operation:.....	7
1.3.2 PV System Structure.....	8
1.3.3 PV Characteristics.....	9
1.4 Architecture of a Solar Power Plant	13
1.4.1 PV Modules	13
1.4.2 Inverters	14
1.4.3 Transformers and Grid Connection	15
1.4.4 Security and Protection.....	17
1.5 Monitoring and Data Acquisition System:	17
1.5.1 Importance of Monitoring.....	17
1.5.2 Data Acquisition and SCADA Systems.....	18
1.6 Environmental Factors Affecting PV Performance	18
1.6.1 Solar Irradiance Variability	18
1.6.2 Temperature Effects.....	18
1.6.3 Wind Cooling Effect.....	19

1.6.4 Humidity and Atmospheric Conditions	19
1.6.5 Soiling and Dust Accumulation	20
1.7 Common Faults in Solar Power Plants	20

CHAPTER II

Fault Detection and Diagnosis Methods in PV Systems

2.1 Introduction.....	24
2.2 Overview of Fault Detection in PV Systems	24
2.2.1 Definition of Fault detection and Diagnosis	24
2.2.2. Objectives of Fault Detection	24
2.2.3 Fault Detection Stages (Detection - Classification - Localization)	25
2.3 Conventional Fault Detection Methods	26
2.3.1 Threshold-Based Methods	26
2.3.2 Rule-Based Techniques	26
2.3.3 Single Processing Approaches.....	27
2.3.4 Limitation of Conventional Methods.....	27
2.4 Model-Based Fault Detection Methods.....	28
2.4.1 Mathematical Modeling of PV Systems	28
2.4.2 Residual Generation Techniques	28
2.4.3 Observer-Based Methods.....	29
2.4.4 Advantages and Disadvantages	29
2.5 Data-Driven Fault Detection Methods.....	30
2.5.1 Statistical-Based Methods.....	30
2.5.2 Comparison of Fault Detection Techniques.....	31
2.5.3 Principal Component Analysis (PCA).....	31
2.5.4 Dynamic Principal Component Analysis (DPCA)	33
2.5.5 Statistical Monitoring Indices	34

CHAPTER III

IMPLEMENTATION OF THE PROPOSED MONITORING FRAMEWORK

3.2 Description of the Oued Kebarit Photovoltaic Station:	39
3.2.1 Site Description.....	39
3.2.2 Dataset Description.....	39
3.3 Proposed Fault Detection and Isolation FDI Architecture	40
3.3.1 Data Acquisition Layer	41
3.3.2 Data Preprocessing Layer	42
3.3.3 PCA Modeling Layer	42

3.3.4 Fault Detection and Localization Layer.....	42
3.4 Practical Implementation and Subspace Parameterization	43
3.4.1 Data Standardization Implementation	43
3.4.2 Static PCA Subspace Partitioning.....	43
3.4.3 Dynamic PCA Spatio-Temporal Expansion	44
3.5 Fault Detection Monitoring Limits and Performance Metrics.....	44
3.5.1 Parameterized Control Chart Thresholds.....	44
3.5.2 Industrial Evaluation Metrics.....	45
3.6 Experimental Results and Discussion	45
3.6.1 Normal Operating Baseline and Information Retainment	45
3.6.2 Physically-Consistent Fault Injection Scenario	47
3.6.3 Comparative Fault Detection and Statistical Threshold Analysis	47
3.6.4 Diagnostic Fault Localization and Variable Isolation	50
3.6.5 Global Classification Performance and Benchmarking Summary	52
3.7 Conclusion:	55
GENERAL CONCLUSION	57
References.....	58

List of Tables

Table 1.1: Classification and characteristics of common sensor fault types.....	22
Table 2.1: Theoretical Fault Isolation and Diagnostic Theoretical Matrix for the PV System.....	31
Table 3.1: Monitored variables of the Oued Kebarit PV installation. monitored variables.....	40
Table 3.2: Programmatic properties of the injected hybrid fault	47
Table 3.3: Quantitative statistical fault detection and latency metrics comparison	49
Table 3.4: Multi-variable fault localization index and diagnostic ranking comparison.....	51
Table 3.5: Definitive global classification and performance matrix benchmarking	53

List of figures

Figure 1.1: Off-Grid Solar Photovoltaic Systems	4
Figure 1.2: Schematic Diagram of an On-Grid Solar Photovoltaic System.....	4
Figure 1.3 : Architecture of a hybrid photovoltaic system	5
Figure 1.4: A p-n junction PV cell	8
Figure 1.5: Theoretical I-V and P-V curves.....	11
Figure 1.6: The ideal model of solar cell	12
Figure 1.7: single diode model of a solar cell	13
Figure 1.8: Solar Panel	14
Figure 1.9: Grid-Connected PV System.....	15
Figure 1.10: PV System with Microinverters.....	16
Figure 1.11: Grid Connection of a Photovoltaic System through a Power Transformer	16
Figure 1.12: PV system with protective diodes.....	17
Figure 1.13 : Influence of irradiance on I-V characteristics of a PV module	18
Figure 1.14 : Influence of Irradiance on the I-V Characteristic	19
Figure 3.1: Flowchart of the Proposed Static and Dynamic PCA-Based FDI Framework for the PV System.	41
Figure 3.2: Explained variance and cumulative percent variance by each principal component. (MATLAB simulation results).....	43
Figure 3.3: DPCA Hotelling T ² Before Fault Injection (MATLAB simulation results).....	46
Figure 3.4: Monitoring Charts for the PV System After Fault Injection	48
Figure 3.5: Comparative Diagnostic Variable Ranking Matrix (MATLAB simulation results) .	51

List of Abbreviations

AC : Alternating Current

DC : Direct Current

DPCA : Dynamic Principal Component Analysis

FAR : False Alarm Rate

FDI : Fault Detection and Isolation

FDR : Fault Detection Rate

I-V : Current-Voltage Curve

MPPT : Maximum Power Point Tracking

P-V :Power-Voltage Curve

PCA :Principal Component Analysis

PV : Photovoltaic

RBC :Reconstruction-Based Contribution

SPE :Squared Prediction Error

STC :Standard Test Conditions

T²: Hotelling's T²-squared Statistic

List of Notations & Symbols

I_{ph} : Photocurrent {A}

I_d: Diode Current {A}

I_s : Diode Saturation Current {A}

I_{sc} :Short-Circuit Current {A}

V_{oc}: Open-Circuit Voltage {V}

V_{th} : Thermal Voltage {V}

P_{max} : Maximum Power Output {W}

R_s: Series Resistance (Ω)

R_{sh}: Shunt Resistance (Ω)

a : Diode Ideality Factor

N_s : Number of Solar Cells Connected in Series

G : Solar Irradiance W/m²

T : PV Cell Temperature °C

μ : Mean Vector of the Training Dataset

σ : Standard Deviation Vector of the Training Dataset

k: Boltzmann constant (1.380649×10^{-23} J/K)

q: Elementary charge ($1.60217662 \times 10^{-19}$ C)

General

Introduction

General Introduction

The engine of industrial advancement and sustainable development is energy. The shift to clean and renewable energy sources has become essential due to the ongoing increase in the world's energy consumption and the growing severity of environmental issues. Photovoltaic (PV) solar energy is one of the most plentiful and sustainable of these sources. Algeria has strategically prioritised the installation of photovoltaic power plants to take use of its remarkable solar irradiation potential and to foster the growth of a green economy.

Photovoltaic power plants have many benefits, but they also have a lot of operating difficulties. Partial shading, short-circuit faults, and sensor deviations are examples of unforeseen failures that can cause significant losses in energy production and hasten equipment deterioration. As a result, Fault Detection and Diagnosis (FDD) has emerged as a crucial prerequisite for guaranteeing the dependability and effectiveness of solar systems, lowering maintenance expenses, and preserving a steady energy supply by early detection of anomalous operating situations.

This dissertation's main research question can be stated as follows in this context: In order to achieve precise and early fault detection and isolation in photovoltaic systems operating under highly variable desert conditions, like those found at the Oued El Kebrit photovoltaic power plant in souk ahras, Algeria, how can an advanced multivariate statistical monitoring model be developed based on a comparative analysis between conventional static Principal Component Analysis (PCA) and its dynamic extension, Dynamic Principal Component Analysis (DPCA)? More precisely, in comparison to traditional static techniques, how much does the inclusion of temporal dynamics contribute to lowering the False Alarm Rate (FAR), increasing the Fault Detection Rate (FDR), and minimising Detection Delay?

The modelling of the photovoltaic system and the effect of faults on its performance, the creation of statistical control limits for the monitoring indices Hotelling's T^2 and Squared Prediction Error (SPE), and the efficiency of Reconstruction-Based Contribution (RBC) charts in precisely identifying and isolating faults within the monitored process are some of the ancillary questions that arise from this primary research problem.

The main goal of this work is to improve problem identification and isolation performance by using data-driven Multivariate Statistical Process Control (MSPC) methodologies. The study specifically focuses on performing a thorough quantitative and qualitative comparison within the MATLAB environment between the dynamic DPCA methodology and the traditional static PCA strategy. Based on actual operational and environmental data gathered from the Oued El Kebrit photovoltaic power plant in Souk ahras, the study aims to assess how well the suggested

methodology improves monitoring performance, increases diagnostic accuracy, and lowers both technical and financial losses related to system malfunctions.

CHAPTER I

Fundamentals of Solar Photovoltaic Systems

1.1 Introduction

The transition to renewable energy sources has accelerated due to growing global energy demand and environmental concerns associated with fossil fuels. Solar energy is regarded as one of the most promising sources due to its availability, sustainability, and minimal environmental impact. Solar photovoltaic (PV) systems convert sunlight directly into electricity using semiconductor materials. Recent technological advances have increased system efficiency while reducing costs, making PV systems widely accessible for both grid-connected and standalone applications. In countries with high solar potential, such as Algeria, photovoltaic systems play a crucial role in maintaining energy security and promoting sustainable development while helping to reduce greenhouse gas emissions [1].

This chapter provides the necessary theoretical foundation for understanding photovoltaic system operation, including its components, environmental sensitivities, and common failure modes. This foundation is essential for developing the data-driven fault detection framework presented in subsequent chapters [1].

1.2 Overview of Solar Photovoltaic Systems

Since the development of the first silicon-based solar cell at Bell Laboratories in 1954, PV technology has evolved into a cost-competitive solution widely integrated into national power grids.

1.2.1 Types of Solar Power Systems:

a. PV systems operate mainly in three basic configurations:

- **Off-Grid Solar Photovoltaic Systems:**

These systems operate independently of the national electricity network. They are classified into two types:

- Direct-coupled systems: Utilize solar power instantaneously as it is produced.
- Battery-integrated systems: Incorporate storage units for continuous power supply during nighttime or low-irradiance conditions.

Standalone systems are particularly effective for decentralized applications such as telecommunication relay stations, remote lighting, and rural infrastructure in regions where the central power grid is absent or unstable.

The fundamental architecture of this configuration is shown in Figure 1.1.

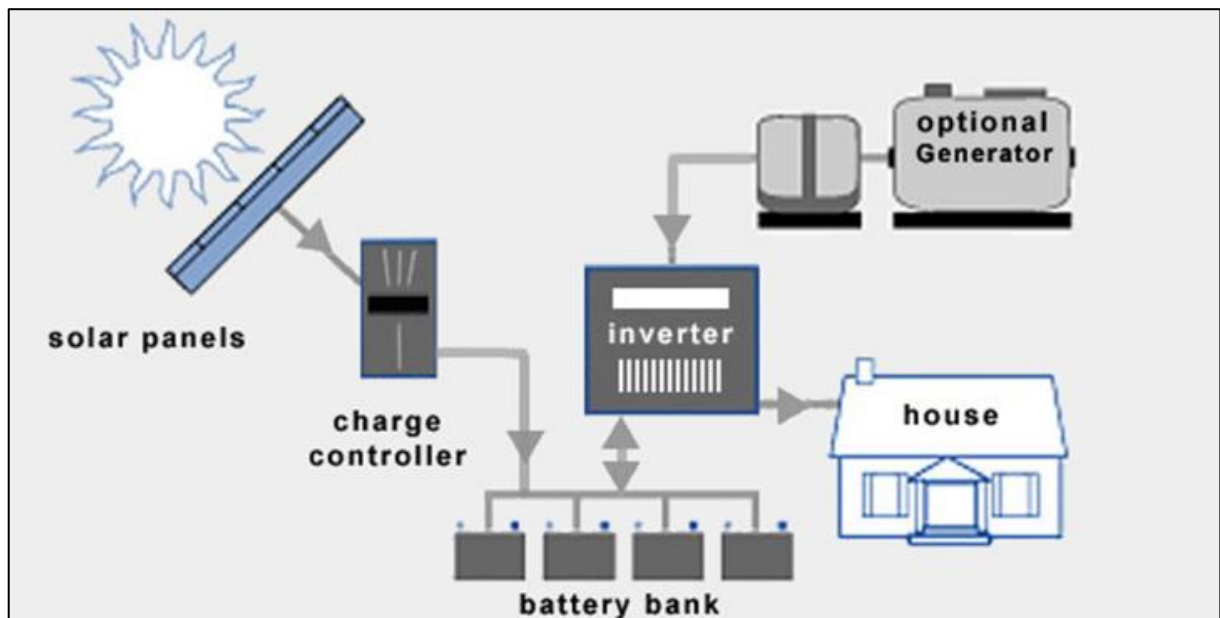


Figure 1.1: Off-Grid Solar Photovoltaic Systems

- **b. Grid-Connected Photovoltaic Systems (On-Grid Systems):**

These systems connect to the public electrical grid via a high-quality inverter that converts DC power from solar panels into AC power that meets grid specifications. During daylight hours, the generated electricity is either used immediately or sold to utility companies. At night, when the system cannot produce electricity, power can be purchased from the grid.

This configuration has two categories: direct grid connection and grid connection with battery backup.

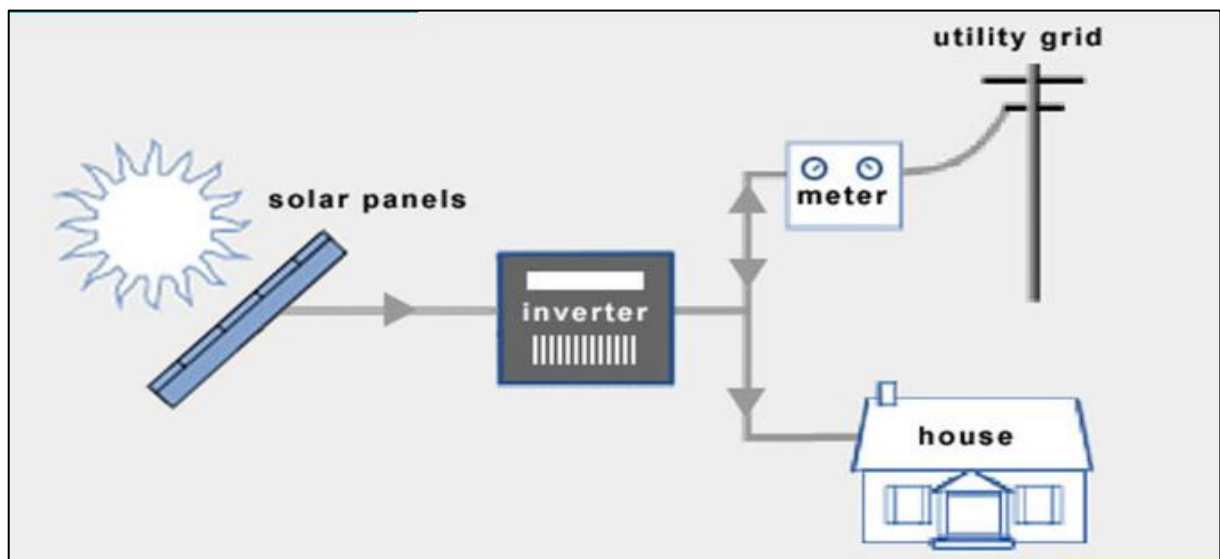


Figure 1.2: Schematic Diagram of an On-Grid Solar Photovoltaic System

- **Hybrid Photovoltaic Systems:**

Hybrid systems combine the advantages of both grid-connected and off-grid configurations. They are connected to the public grid and include a battery storage system, allowing energy storage for nighttime use or during power outages. These systems are particularly suitable for residences where the public grid is erratic or unreliable.

Alternatively, hybrid systems may integrate diesel generators alongside photovoltaic energy, making battery storage optional (as illustrated in Figure 1.3).

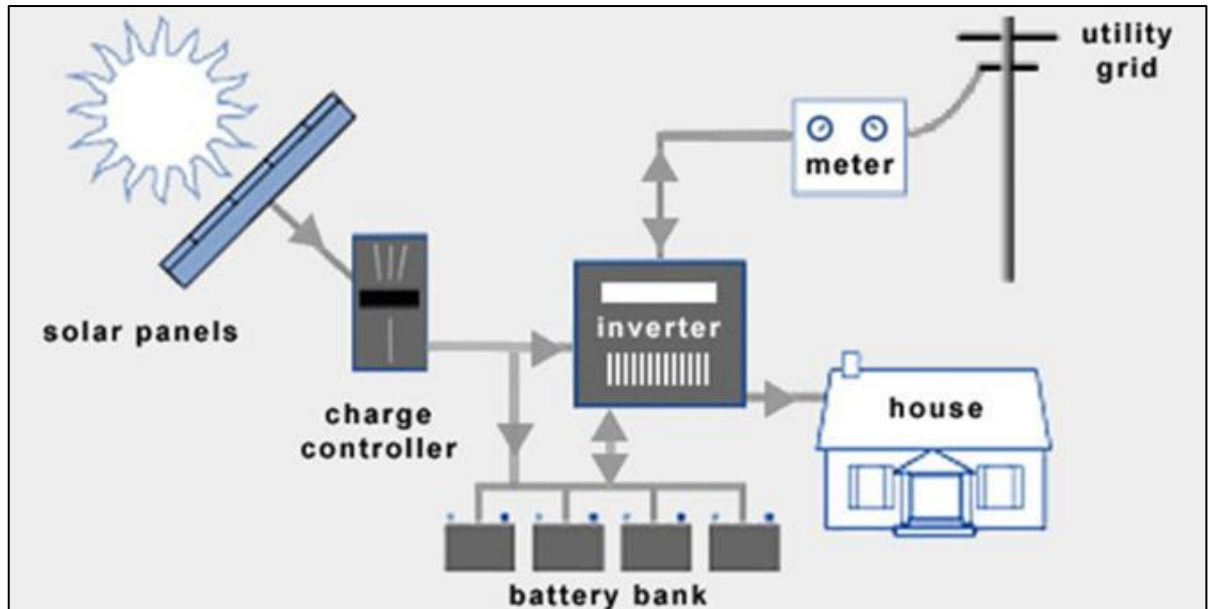


Figure 1.3 : Architecture of a hybrid photovoltaic system [2]

1.2.2 Advantages and Challenges of PV Systems

a. Advantages:

- **High reliability:**

Despite being exposed to harsh weather conditions, photovoltaic systems remain highly reliable. Using PV arrays, energy sources can operate continuously without interruption.

- **Long operational lifespan:**

Solar panels are usually covered by warranties of 25 years or more, and most of them continue operating for many additional years.

- **Clean energy source:**

PV cells provide clean and green energy, which is their most important advantage. There is no concern about harmful greenhouse gas emissions such as carbon dioxide into the atmosphere.

- **Free raw energy source:**

It is also beneficial because no raw materials need to be purchased. Solar cells rely on the sun to generate electricity, which is abundantly available. Choosing solar energy is an excellent option. The system requires an initial investment; however, it is natural, free in operation, and cost-saving in the long term. Once PV-generated energy is used, electricity costs are significantly reduced.

- **Low Maintenance Costs:**

One of the most significant advantages of photovoltaic systems is their minimal maintenance requirements. Unlike conventional thermal power systems that rely on fuel and contain numerous moving parts, PV systems are characterized by their structural simplicity. Consequently, they only require routine inspections and periodic cleaning to maintain efficiency, leading to a substantial reduction in long-term operational expenses.[3]

b. Challenges

- **Variability and Intermittency:** One of the primary technical hurdles in grid integration is the intermittent nature of solar energy. Power generation is inherently fluctuating due to diurnal cycles, seasonal changes, and unpredictable atmospheric conditions. This variability often creates a significant gap between peak production and consumer demand, which can compromise the overall stability of the electrical grid.
- **Forecasting and Predictive Techniques:** To mitigate uncertainty, advanced forecasting models are increasingly employed, leveraging Artificial Intelligence (AI) and Machine Learning (ML) techniques. While algorithms such as Genetic Algorithms (GA) and Support Vector Machines (SVM) have improved prediction accuracy, capturing the complexity of sudden weather shifts and seasonal transitions remains a critical area for further research and optimization.
- **Voltage and Frequency Regulation:** High penetration of PV systems can introduce significant voltage fluctuations within the network. Since solar inverters are responsible for converting DC to AC, their ability to provide reactive power support and maintain stable

voltage levels is crucial. Inadequate regulation by these inverters can lead to power quality issues, especially in localized distribution networks.

- **Grid Infrastructure and Upgrades:** The existing electrical infrastructure often faces strain when integrating large-scale PV power flows. To accommodate these bidirectional and variable loads, substantial upgrades are necessary, including the reinforcement of transmission lines and the modernization of substations. While implementing "Smart Grid" management technologies is essential, these modifications involve significant capital investment and complex logistical planning. [4]

1.3 Photovoltaic Fundamentals

1.3.1 PV Cell Operation:

According to the international standards established by IEC 61836 , a photovoltaic (PV) system is defined as a setup that directly converts solar radiation into electrical energy. This conversion process relies on the photoelectric effect, which typically occurs within semiconductor materials, most notably Silicon (Si).

Cell Structure and Composition:

A standard silicon solar cell is engineered using two distinct layers to create a junction:

- a. The n-layer (Negative):** A thin top layer heavily doped with elements like Phosphorus, creating a high concentration of free electrons.
- b. The p-layer (Positive):** A thicker base layer lightly doped with elements such as Boron, resulting in a deficiency of electrons (creating "holes").
- c. The Photovoltaic Effect:** The operation of the cell can be described through the following physical stages:
 - c. Photon Absorption:** When incident sunlight strikes the n-layer, photons with energy exceeding the silicon bandgap are absorbed. This energy breaks the covalent bonds within the silicon crystal, generating electron-hole pairs.
 - d. Charge Separation:** An internal electric field exists at the p-n junction (the depletion region). This field acts as a barrier that drives free electrons toward the n-side and holes toward the p-side, preventing immediate recombination and establishing a potential difference (open-circuit voltage).
 - e. Current Generation:** Once an external load is connected, this potential difference forces electrons to flow from the n-layer through the external circuit toward the p-

layer. This continuous movement of charges constitutes a Direct Current (DC), providing usable electrical power

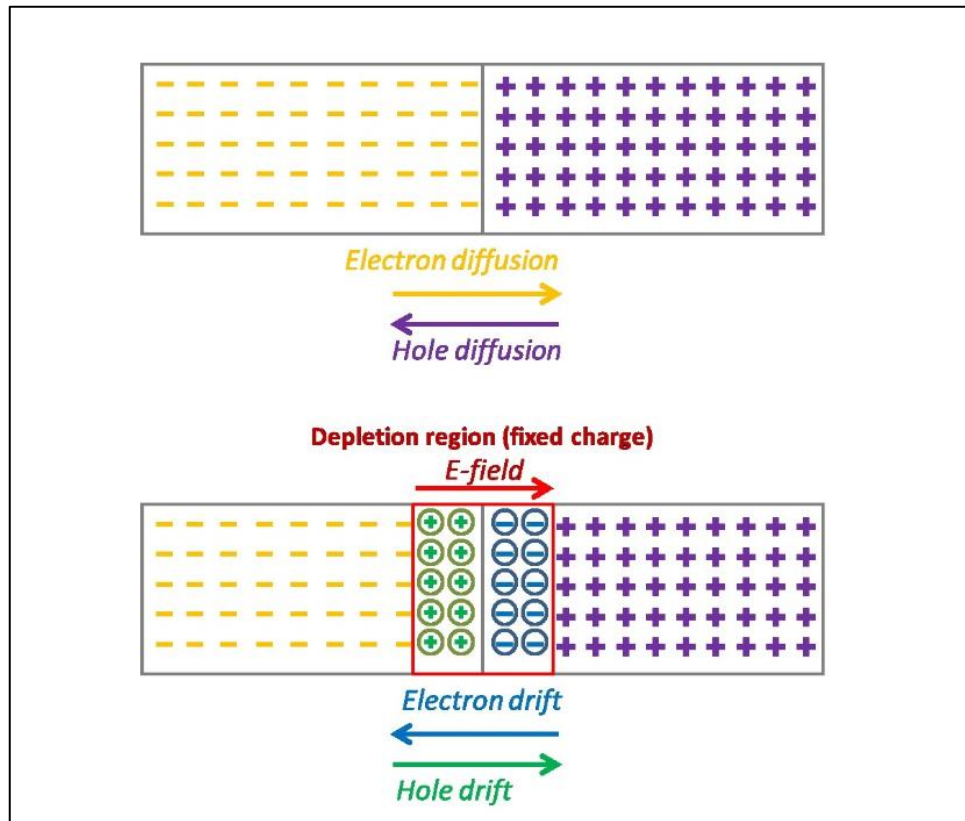


Figure 1.4: A p-n junction PV cell [5]

1.3.2 PV System Structure

A complete PV system consists of several key components working together:

a. Solar Panels:

The core component that converts solar radiation into DC electricity. Panels are formed by connecting multiple cells in series to achieve the desired voltage [6].

b. The Solar Inverter:

The inverter is an essential power electronics device that converts the Direct Current (DC) output from the PV panels into Alternating Current (AC). This conversion is necessary to ensure the electricity is compatible with standard appliances and grid requirements. Additionally, modern inverters are vital for monitoring system performance and ensuring power stability.

c. Mounting Structure:

The mounting structure is the mechanical framework designed to secure PV panels onto rooftops or ground-based installations. Beyond providing structural stability, its primary role is to

fix the panels at an optimal tilt angle and orientation. This optimization is crucial for maximizing solar irradiance capture and ensuring the overall efficiency of the system.

d. Racking and BOS Components:

The Balance of System (BOS) encompasses all auxiliary components required for the integration and operation of the PV installation, excluding the panels and inverters. This includes electrical wiring, racking, fuses, and switching devices. These elements provide the necessary infrastructure for power distribution and structural support, ensuring the system functions as a cohesive and safe electrical unit.

e. Monitoring System:

The monitoring system tracks and analyzes the real-time performance of the PV installation. By collecting data on energy yield and efficiency, it enables the early detection of faults and performance degradation. This ensures the system operates at its maximum potential and remains reliable [6].

1.3.3 PV Characteristics

Understanding PV system performance requires familiarity with the following key characteristics:

a. PV Characteristics:

The characteristics of a PV system are essential for understanding its performance and behavior. The key characteristics include:

- **I–V Curve:**

The I–V curve illustrates the electrical relationship between the output current and voltage of a PV system. It is a critical diagnostic tool used to determine the Maximum Power Point (MPP), ensuring the system operates at its peak power capacity under given conditions.

- **P–V Curve:**

The P–V curve illustrates the relationship between power output and voltage. It is used to identify the Maximum Power Point (MPP), representing the specific voltage level where the system achieves its peak power generation.

- **Maximum Power Point (MPP):**

The unique operating state on both I-V and P-V curves where the system delivers its maximum power output.

- **Open-Circuit Voltage (Voc):**

The **Voc** is the maximum voltage a PV system can achieve when the external circuit is open and the current is zero. It represents the upper electrical limit of the system and is a key parameter for defining the characteristic I–V curve under specific environmental conditions.

- **Short-Circuit Current (Isc):**

The **Short-Circuit Current** represents the maximum current flow through the PV system when the load resistance is zero (terminals are shorted). Beyond this parameter, the overall performance and conversion efficiency (τ) of the cell are evaluated using the following variables:

- **Pmax:** The peak power output achieved at the MPP.
- **E:** The incident solar irradiance (W/m^2) under Standard Test Conditions (STC).
- **Ac:** The effective surface area of the PV cell (m^2).

System Efficiency and Evaluation:

The efficiency of a photovoltaic cell is not a fixed value; it is highly dependent on the semiconductor material, manufacturing precision, and real-world operating conditions specifically **irradiance levels** and **cell temperature**. Analyzing these characteristics is fundamental for the design, optimization, and comparative assessment of different PV technologies, as they directly dictate the system's ability to maximize energy yield.

- **Fill Factor (FF):**

The Fill Factor (FF) measures the quality and "squareness" of the I–V curve. It represents the ratio between the actual maximum power and the theoretical power limit higher FF indicates minimal internal losses and superior cell performance, making it a vital indicator for detecting efficiency drops and system degradation.

$$FF = \frac{V_{mp} I_{mp}}{V_{oc} I_{sc}} \quad \text{I.1}$$

I_{mp}: Current at the Maximum Power Point [A]

V_{mp}: Voltage at the Maximum Power Point [V]

- **Efficiency :**

The percentage of incident solar energy converted into electrical power:

$$\eta = \frac{P_{max}}{E \cdot A_c} \times 100 \quad \text{I.2}$$

Where E is the incident solar irradiance (W/m²) under Standard Test Conditions (STC) and A_c is the effective cell area (m²).

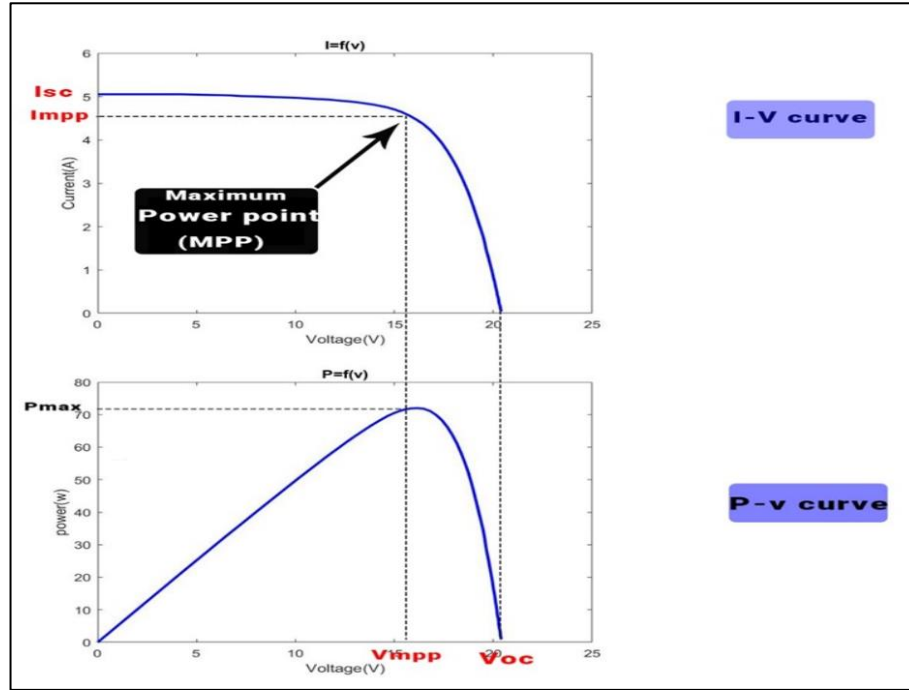


Figure 1.5: Theoretical I-V and P-V curves [5]

1.3.4 Mathematical Modeling of PV Cells

Accurate PV device modeling is essential for predicting system performance under varying operating conditions. Among the various models available, the single-diode model strikes an optimal balance between accuracy and simplicity, making it the most widely used in PV system analysis .

a. Ideal Cell Modeling:

The ideal model assumes a lossless PV device with no series or shunt resistances. While suitable for basic analysis, it is insufficient for real-world applications. The output current is given by:

$$I = I_{ph} - I_0 \left(e^{\frac{q(V+RsI)}{AkT}} - 1 \right) \quad \text{I.3}$$

Where:

- I : Output current [A]
- I_{ph} : Photo generated current [A]
- I_0 : Dark saturation current [A]
- V : Output voltage [V]
- q : Elementary charge ($1.60217662 \times 10^{-19}$ C)
- k : Boltzmann constant (1.380649×10^{-23} J/K)
- T : Ambient temperature [K]
- A : Diode ideality factor

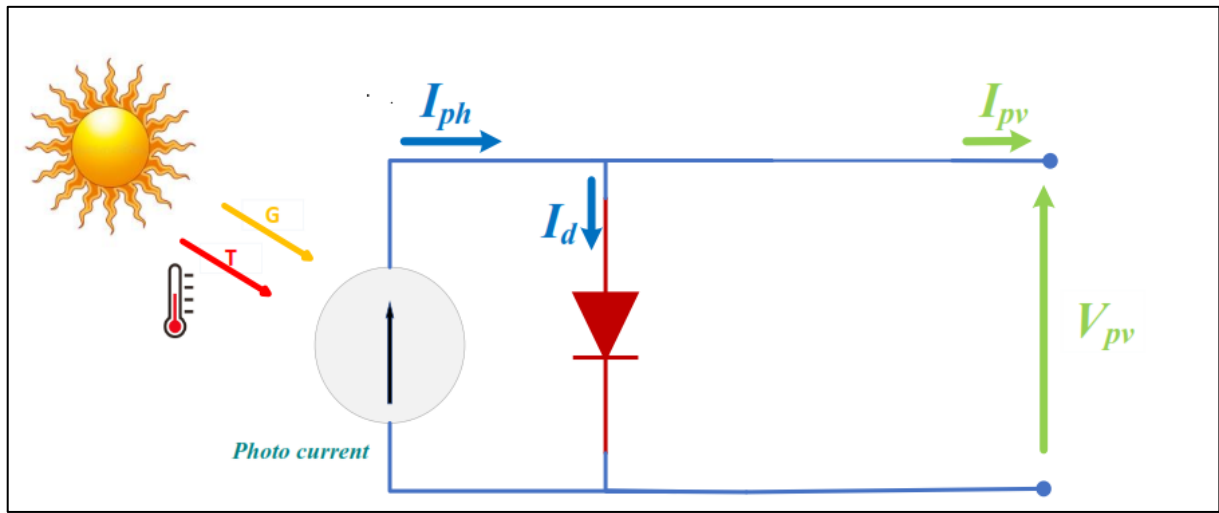


Figure 1.6: The ideal model of solar cell

b. Single-Diode Model:

A more accurate and realistic depiction of PV behavior is introduced by the single-diode model (Figure 1.6). It takes into consideration elements including shade, temperature, and cell deterioration. It models the I-V curve using a single diode.

However, certain factors that are frequently missing from manufacturer datasheets are necessary for correct modeling. While other parameters like R_s , R_{sh} , A , I_{ph} , and I_0 must be approximated using parameter extraction techniques, V_{oc} , I_{sc} , I_{mpp} , and V_{mpp} are typically available.

The output current is defined as:

$$I = I_{ph} - I_0 \left(e^{\frac{q(V+R_s I)}{A k T}} - 1 \right) - \frac{V + R_s I}{R_{sh}} \quad \text{I.4}$$

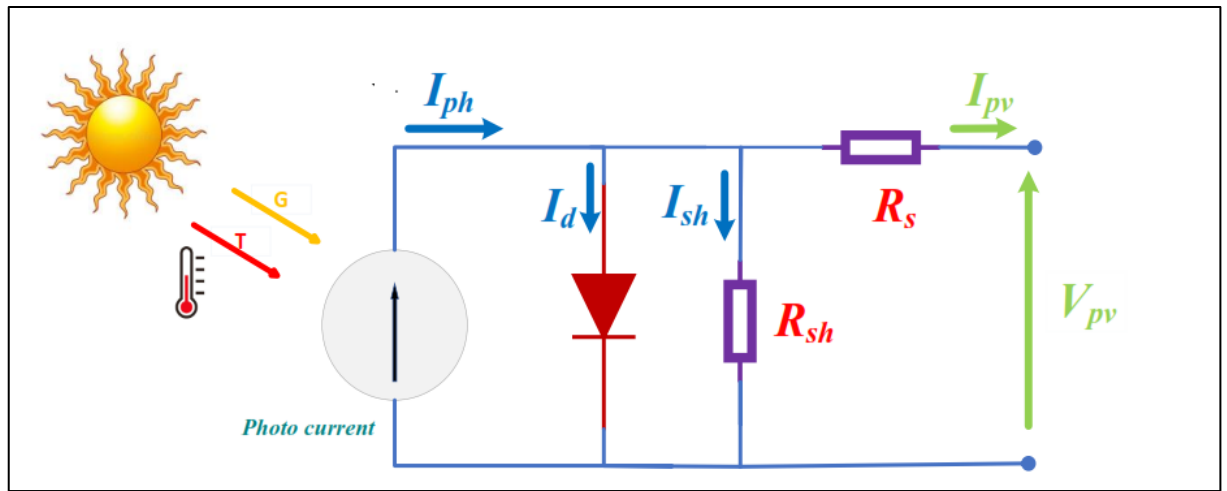


Figure 1.7: single diode model of a solar cell

Where:

- R_{sh} : Shunt resistance (Ω)
- Other parameters are as defined above

1.4 Architecture of a Solar Power Plant

1.4.1 PV Modules

Solar panels consist of photovoltaic cells connected in series to achieve the required voltage. Several interconnected cells form a module, which is the smallest power-generating unit that can be assembled and dismantled. Multiple modules are arranged into arrays and connected according to system specifications to create photovoltaic fields. Depending on current technology, cells are typically made of crystalline silicon (monocrystalline or polycrystalline) in the form of wafers, strips, or thin layers. To increase power output, modules are connected in both parallel and series configurations.

Each solar cell is protected by a bypass diode connected in parallel to prevent overheating and short circuits. Additionally, an anti-reverse current diode is installed in each branch to prevent the solar module from acting as a receiver rather than a generator. [7]

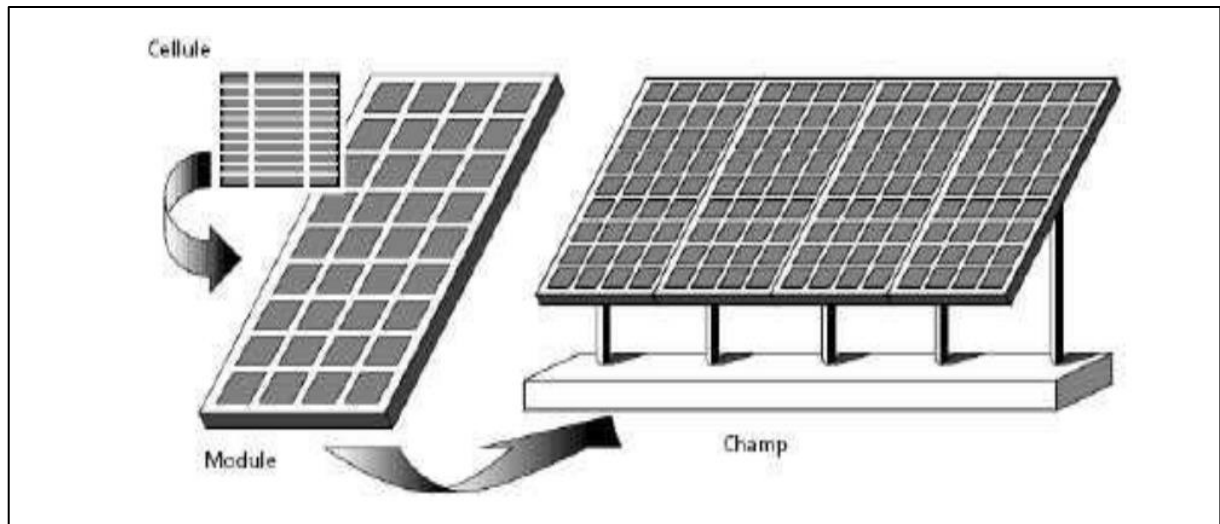


Figure 1.8: Solar Panel [7]

1.4.2 Inverters

Direct current (DC), which includes current, voltage, and power, is a fundamental feature of photovoltaic outputs. However, the norm for producing, transmitting, and distributing electricity is alternating current (AC). As a result, inverters which transform DC output into usable AC electricity are an essential part of PV systems.

An electrical circuit or device that changes direct current (DC) into alternating current (AC) is called an inverter. The topology and design of the device determine the input voltage, output voltage, frequency, and total power.

Inverters have a major impact on the efficiency and dependability of solar systems, making them one of the most crucial parts. They transform PV array electricity into usable AC power. Furthermore, contemporary inverters include sophisticated features like data monitoring, anti-islanding prevention, and Maximum Power Point Tracking (MPPT).

a) Types of Inverters:

- **Standalone inverters:**

Designed for off-grid applications or remote systems with battery backup, providing stable voltage and frequency for electrical loads.

- **Grid-tied inverters:**

The most commonly used type, designed to automatically shut down when grid power is lost for safety reasons.

- **Hybrid inverters:**

Designed for hybrid energy systems combining PV arrays with other renewable sources such as wind turbines. These inverters can operate both as standalone and grid-connected systems. [4]

1.4.3 Transformers and Grid Connection

a) Transformer for strings :

In photovoltaic systems, a string transformer is used to change DC electricity from solar panels into AC electricity that may be used in homes or businesses. A string arrangement of solar panels is referred to as a "string."

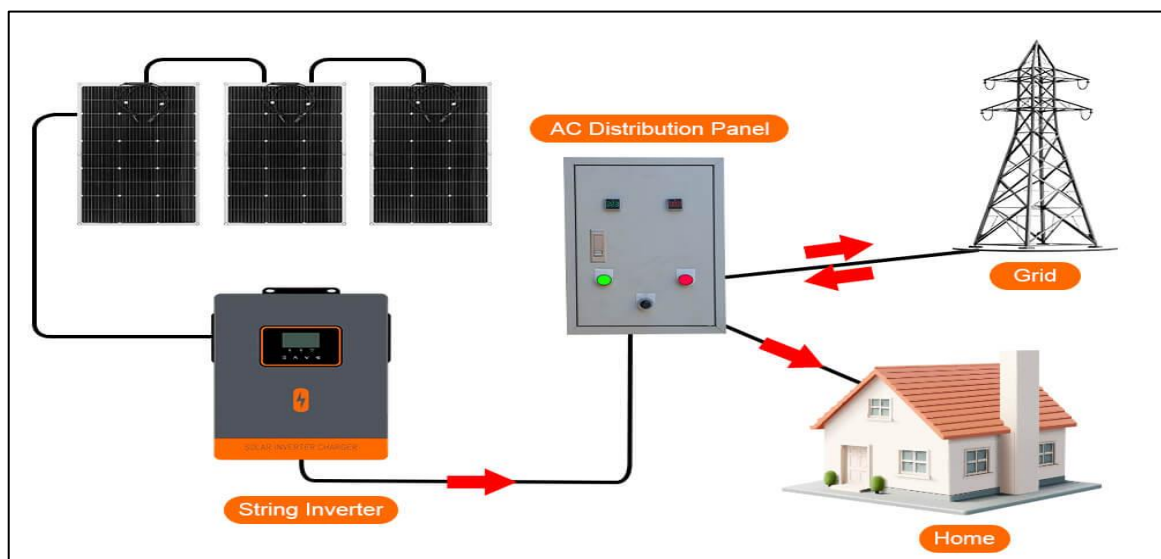


Figure 1.9: Grid-Connected PV System

b) The microinverter :

A microinverter is a tiny, standalone inverter that is fastened to the rear of every solar panel. It allows each panel to function independently by converting DC current from a single panel into AC electricity.

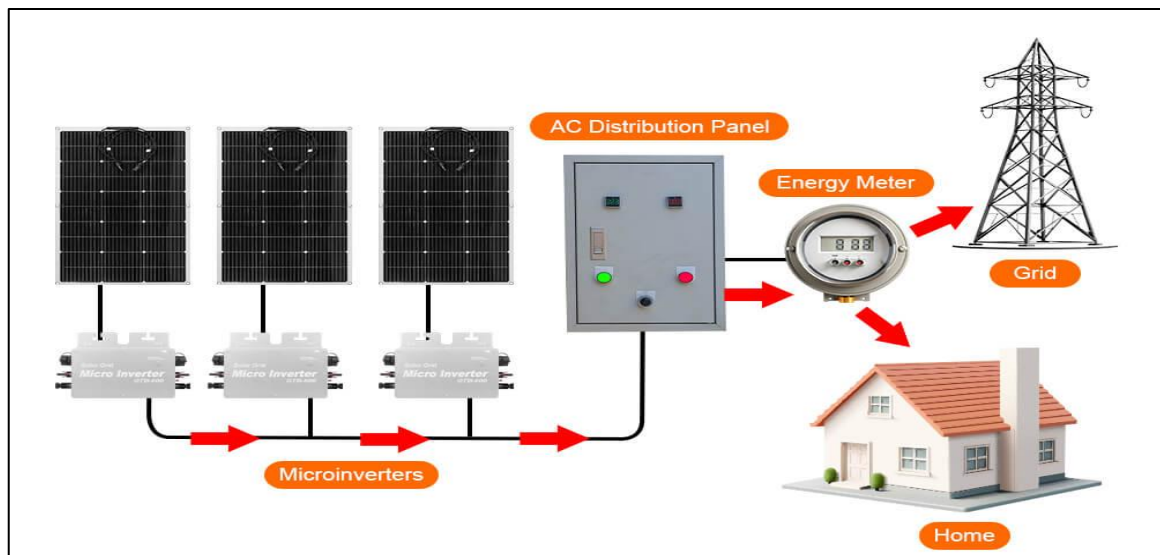


Figure 1.10: PV System with Microinverters

c) Grid Connection:

Solar systems are connected to the main electrical grid using power transformers. In order to guarantee efficient energy flow and equipment safety, they modify the voltage levels between solar panels and the grid.

Voltage levels produced by solar energy frequently fall short of grid needs. Transformers ensure reliable and safe energy delivery by modifying voltage levels to comply with grid regulations [8].

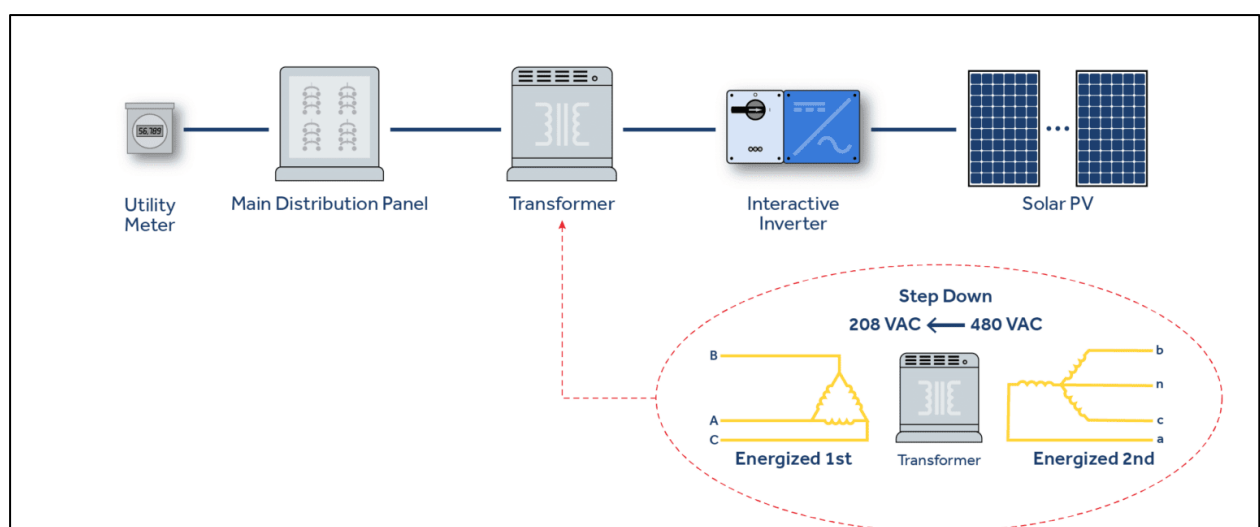


Figure 1.11: Grid Connection of a Photovoltaic System through a Power Transformer [2].

1.4.4 Security and Protection

Electrical protection is crucial for increasing system longevity and preventing failures caused by cell connectivity issues or shading. Protection diodes prevent a damaged cell from impairing overall performance. Two types are used:

- Series diodes: Provide anti-backflow protection, stopping reverse current.
- Bypass diodes: Prevent energy losses by rerouting current around darkened cells.

A typical solar panel contains one to three bypass diodes (approximately 36 cells per 3 diodes) [5].

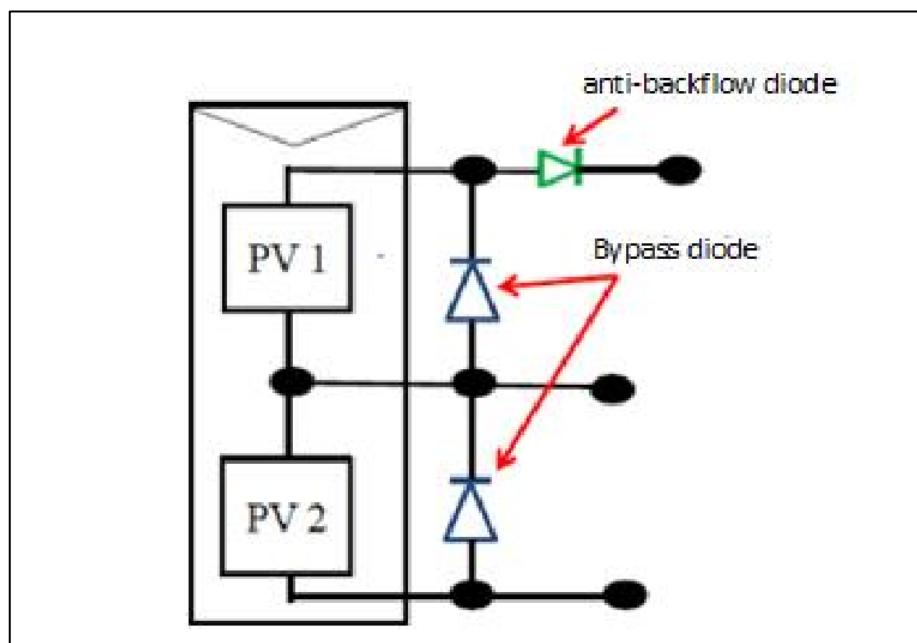


Figure 1.12: PV system with protective diodes [5].

1.5 Monitoring and Data Acquisition System:

1.5.1 Importance of Monitoring

The installation of a dependable monitoring system is crucial to guaranteeing the economical and efficient functioning of photovoltaic power generation systems and optimizing their potential. This system is essential for monitoring and evaluating the performance of PV systems, minimizing power outages, and guaranteeing seamless operation. It also makes it possible to evaluate performance using a variety of techniques. Utilizing monitoring systems promotes PV energy systems' long-term operation and best use [9].

1.5.2 Data Acquisition and SCADA Systems

Electrical generation and system performance can be monitored using a variety of methods.

A reliable monitoring strategy is represented by Supervisory Control and Data Acquisition (SCADA) systems. SCADA systems can be used for real-time data collecting, processing, control, and remote monitoring. They are crucial to contemporary photovoltaic energy systems because of their clever data handling skills [10].

1.6 Environmental Factors Affecting PV Performance

1.6.1 Solar Irradiance Variability

Under clear skies, standard solar irradiance conditions (also known as "one sun" or "peak sun") are assumed to be 1000 W/m^2 . Photovoltaic panel current production is proportionally reduced below this value. For example, under half-sun conditions (500 W/m^2), the output current is reduced by approximately half.

Standard Test Conditions (STC) are defined as 1000 W/m^2 irradiance, 25°C cell temperature, and air mass 1.13. Under these conditions, current varies significantly with irradiance, while voltage increases only slightly

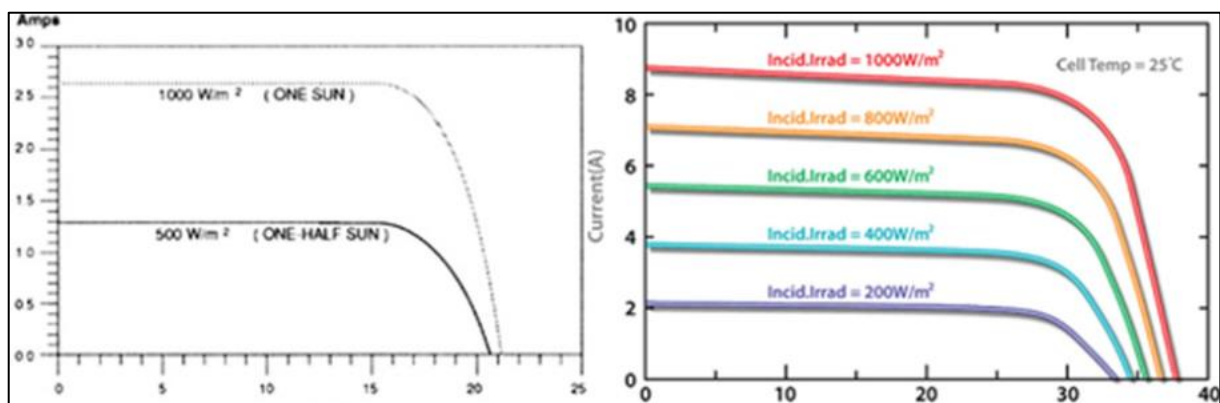


Figure 1.13 : Influence of irradiance on I-V characteristics of a PV module

1.6.2 Temperature Effects

As illustrated in Figure (1.13), a rise in solar panel temperature causes a voltage drop of roughly 0.04 to 0.1 volts for every 1°C increase in temperature (-0.04 to $-0.1 \text{ V/}^\circ\text{C}$). In order to avoid overheating and performance loss, solar panels should not be placed directly on rooftops; instead, airflow must be permitted behind each panel. To guarantee adequate cooling, a ventilation

gap of around 4 to 6 inches is usually necessary. The inverse relationship between PV panel temperature and voltage output is shown in Figure (1.14).

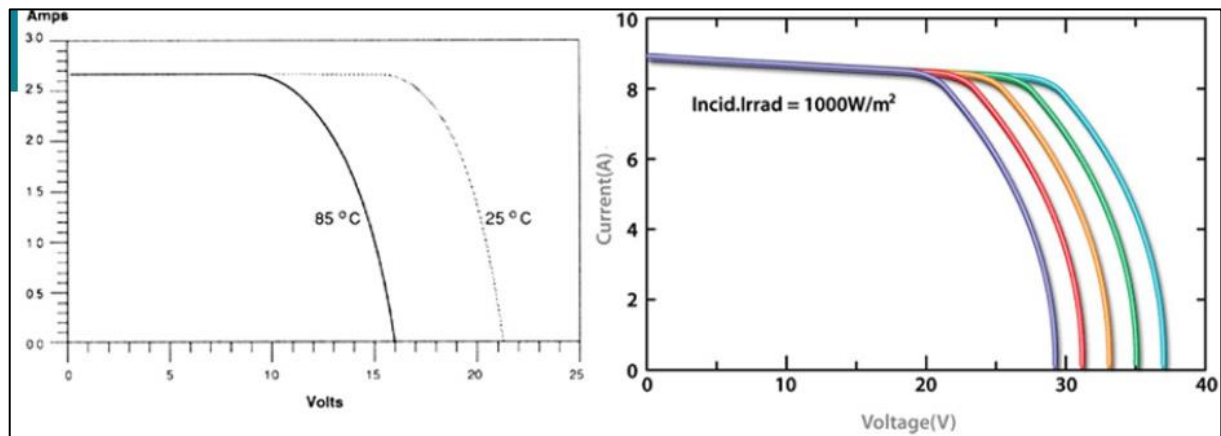


Figure 1.14 : Influence of Irradiance on the I-V Characteristic [11]

1.6.3 Wind Cooling Effect

The temperature distribution of solar panels is greatly influenced by wind direction and speed during actual working conditions, which in turn has an impact on system efficiency. This has prompted a number of research that concentrate on the temperature behavior of PV modules in actual outdoor wind conditions.

In their study of PV module temperature under various environmental circumstances, Gaszczor *et al.* emphasized the significance of wind direction and speed. According to Jokman *et al.* annual energy production is underestimated by about 3.5% when wind effects are ignored in PV calculations [9].

1.6.4 Humidity and Atmospheric Conditions

There are two primary ways that humidity can lower PV efficiency:

- Water vapor or tiny water droplets may build up on solar panels, deflecting or refracting sunlight away from the solar cells and lowering incident solar radiation and power production.
- Solar panels may deteriorate after prolonged exposure to hot, muggy weather. Both thin-film and crystalline silicon technologies are impacted by this. In tropical regions, however, cadmium telluride (thin-film) solar cells perform roughly 5% better [12].

1.6.5 Soiling and Dust Accumulation

Dust buildup can drastically lower energy production. Dust buildup on PV panel surfaces creates a coating that prevents sunlight from reaching the solar cells, which lowers energy output. Particles carried by the wind settle on panel surfaces, causing a slow buildup of dust. This process can happen quickly in areas with a lot of dust, resulting in large efficiency losses [13].

1.7 Common Faults in Solar Power Plants

Photovoltaic power plants are complex systems integrating electrical and environmental components. During operation, various faults can occur, impacting system efficiency and performance. Understanding these fault types is essential for developing effective fault detection and diagnosis strategies.

Faults in photovoltaic systems can be classified into two main categories:

a. System faults:

Electrical and physical faults affecting PV components, including short circuits, open circuits, inverter failures, module degradation, soiling, and shading.

b. Sensor faults:

Failures affecting measurement sensors, including irradiance sensors, temperature sensors, wind speed sensors, humidity sensors, and pressure sensors. These faults are particularly critical for data-driven monitoring approaches since they directly affect the quality of measured data used for fault detection.

This thesis focuses on the detection and diagnosis of sensor faults in PV systems, as these are essential for ensuring reliable data acquisition and accurate system monitoring [2].

1.7.1 System Faults (Brief Overview)

a. Short Circuit:

A short circuit occurs when two points with different electrical potentials become abnormally connected, causing very high current to flow outside the intended path. If not properly managed, this fault may activate protection mechanisms or damage system components.

b. Open Circuit:

An open circuit results from a break in the system's electrical path, completely stopping current flow. Consequently, the affected portion of the system loses energy production.

c. Inverter Faults:

The inverter is a critical system component. Inverter faults include MPPT (Maximum Power Point Tracking) malfunctions, operational shutdown, and failure to convert DC to AC. These issues significantly reduce total energy production.

d. Photovoltaic Module Faults:

Photovoltaic module faults include gradual power output deterioration over time, microcracks in solar cells, and bypass diode failure. These issues cause overall performance and efficiency decline.

e. Soiling:

Soiling refers to the accumulation of dust and dirt on photovoltaic module surfaces, which reduces the solar irradiance reaching the cells and consequently decreases energy production. This problem is particularly prevalent in arid and desert areas, such as the El Oued region where our case study is located.

f. Shading:

Shading occurs when nearby objects (trees, buildings, or other obstacles) cast shadows on portions of photovoltaic modules. This fault significantly impacts electrical string performance and may reduce the output of the entire string, not only the shaded portion.

1.7.2 Sensor Faults (Focus of this Thesis)

Sensor faults are failures or inaccuracies in the measurement devices that collect data from the PV system. Since data-driven fault detection methods rely entirely on the quality of measured data, sensor faults can severely compromise monitoring performance if not detected and isolated promptly.

Sensor faults can take several forms:

- a. **Bias fault (Offset fault):** A constant value added to or subtracted from the true measurement. For example, an irradiance sensor consistently reading 200 W/m² above the actual value.
- b. **Drift fault:** A gradual, progressive deviation from the true value over time. For example, a temperature sensor slowly increasing its reading by 0.03 per sample due to aging or calibration loss.
- c. **Complete failure:** The sensor stops providing valid measurements (e.g., constant output, no signal, or NaN values).
- d. **Scaling error (Gain fault):** The sensor reading is multiplied by an incorrect factor (e.g., wind speed sensor reading twice the actual value) [14].

Table 1.1 summarizes the sensor fault types relevant to this thesis, along with examples from PV monitoring systems.

Table 1.1: Classification and characteristics of common sensor fault types

Fault Type	Description	Example in PV System
Bias fault (Offset)	Constant offset added to true value	Irradiance sensor reading 200 W/m ² above actual
Drift fault	Gradual deviation over time	Temperature sensor increasing by 0.03 per sample
Complete failure	Sensor stops providing valid data	Sensor output constant or NaN
Scaling error (Gain fault)	Incorrect multiplication factor	Wind speed sensor reading 2× actual value

1.8 Chapter Summary

This chapter provided an overview of photovoltaic systems, including their basic operation principles, main components, and environmental factors that influence their performance. The single-diode model was presented as the mathematical basis for PV system characterization.

Additionally, common faults in PV systems were classified into system faults and sensor faults, with sensor faults being the main focus of this work.

The next chapter presents fault detection and diagnosis methods in PV systems, with emphasis on data-driven statistical approaches including Principal Component Analysis (PCA) and Dynamic Principal Component Analysis (DPCA) [15].

CHAPTER II

**Fault Detection and Diagnosis Methods in PV
Systems**

2.1 Introduction

Solar energy is one of the most environmentally friendly and sustainable energy sources. According to the International Energy Agency [16], the installed capacity of photovoltaic solar energy systems has been increasing rapidly worldwide.

However, various factors can affect PV system performance, particularly electrical issues within cells that can negatively impact efficiency. Any malfunction in a PV module leads to changes in system performance, requiring precise monitoring and control [17], and may ultimately result in serious safety concerns [18].

To ensure reliability and efficiency, PV systems must be equipped with monitoring and fault diagnosis capabilities capable of detecting problems in a timely manner. Numerous fault detection and diagnosis methods have been proposed in the literature, requiring high speed, advanced processing, and precise sensing to accurately identify a wide range of faults.

2.2 Overview of Fault Detection in PV Systems

Photovoltaic systems have been widely deployed worldwide thanks to technological advances. However, these systems require continuous monitoring to ensure reliable operation.

PV systems are subject to various fault types, ranging from temporary to permanent failures. Identifying the type and location of a fault quickly and cost-effectively without affecting system efficiency remains a challenge. Therefore, effective fault detection and diagnosis methods are essential to protect PV systems from degradation and failures [19].

2.2.1 Definition of Fault detection and Diagnosis

Fault detection and diagnosis (FDD) is defined as a procedure for identifying and isolating operational flaws within a system, utilizing techniques such as data-driven methods and knowledge-driven methods to analyze fault scenarios and improve system reliability. [20]

2.2.2. Objectives of Fault Detection

a. Ensuring continuous energy production:

Early fault detection helps avoid reductions in electrical power output and maintains continuous system operation.

b. Improving system efficiency:

Faults (especially on the DC or AC side) lead to efficiency losses; therefore, detecting them helps maintain the optimal performance of photovoltaic panels and inverters.

c. Reducing economic losses:

By minimizing downtime and reducing losses caused by decreased energy production or damage to system components.

d. Enhancing safety:

Some faults, such as short circuits or ground faults, may pose serious risks; early detection reduces the likelihood of accidents.

e. Protecting system components:

Rapid fault detection prevents minor issues from developing into permanent damage in panels, cables, or inverters.

f. Improving maintenance strategies:

It supports the transition from traditional periodic maintenance to intelligent (condition-based) maintenance.

g. Accurate fault diagnosis and classification:

The reference emphasizes the importance of identifying the type of fault (internal, external, partial, shading, degradation, etc.) to apply the appropriate solution quickly.

h. Enhancing system reliability and sustainability:

By reducing unexpected failures and improving the operational lifetime of photovoltaic systems. [21]

2.2.3 Fault Detection Stages (Detection - Classification - Localization)

a. Fault Detection:

Continuous monitoring of the photovoltaic system performance is carried out by analyzing electrical variables such as voltage, current, and power, and correlating them with environmental conditions such as solar irradiance and temperature. A fault is detected when a clear deviation is observed between measured values and expected or reference values, indicating abnormal system behavior compared to its normal operating state.

b. Fault Classification:

The nature of the detected fault is identified by analyzing electrical signals and performance characteristics. According to the reference, faults in photovoltaic systems are classified into electrical faults such as short circuits and open circuits, environmental-related faults such as partial shading and dust accumulation, as well as faults related to system components such as panels and inverters. This classification is based on the analysis of current–voltage (I–V) characteristics or by using pattern recognition methods.

c. Fault Localization:

This stage aims to isolate the affected part within the photovoltaic system by analyzing performance at different levels, including the system level, string level, and module level. It relies on comparing the performance of different strings within the same array and analyzing variations in currents and voltages, which enables increasingly precise identification of the faulty region or component responsible for the malfunction. [22]

2.3 Conventional Fault Detection Methods**2.3.1 Threshold-Based Methods**

Threshold-based methods for fault detection in photovoltaic systems rely on a simple principle based on comparing measured system variables with predefined reference limits in order to identify deviations that may indicate performance anomalies. In this context, key variables such as voltage, current, and power are first measured, and then compared with reference values, which may be nominal, derived from a mathematical model, or obtained from historical system data.

After that, the magnitude of the deviation between the measured and expected values is calculated. If this deviation exceeds a certain predefined threshold, it is considered an indication of a fault. This principle can be mathematically expressed by comparing the absolute difference between the measured value and the reference value with a specific limit, where exceeding this limit indicates an abnormal condition in the system.

This approach is widely used in applications such as detecting power losses, identifying string imbalances, and monitoring gradual performance degradation, by assigning specific thresholds to each performance indicator such as power output or performance ratio.

Despite their simplicity and ease of real-time implementation, these methods have several limitations. The most important is their sensitivity to the selection of threshold values and their dependence on varying environmental conditions such as solar irradiance and temperature. This may lead to false alarms or reduced detection accuracy compared to more advanced methods. [23]

2.3.2 Rule-Based Techniques

Rule-based techniques for fault detection in photovoltaic systems rely on comparing the actual system performance with the expected performance under the same operating conditions. First, the expected electrical power output is estimated based on solar irradiance, temperature, and system parameters. Then, the performance ratio (PR) is calculated by dividing the measured power from the inverter by the expected power.

After that, a set of conditional “If–Then” rules is applied to determine the system’s operating state. A decrease in the performance ratio below a predefined threshold is considered an indication of underperformance. Additional operational patterns are also used to distinguish different types of faults. For example, zero power output despite high solar irradiance may indicate inverter tripping, while a constant power output despite increasing irradiance can suggest power clipping. Similarly, a partial and recurring drop in power during specific time periods is often associated with partial shading, whereas a gradual long-term decline in output may indicate degradation or soiling effects.

Thus, this methodology is based on predefined expert rules and fixed thresholds without requiring machine learning techniques, making it simple to implement, easy to interpret, and suitable for real-time monitoring systems with limited data availability. [24]

2.3.3 Single Processing Approaches

The single-signal processing approach in photovoltaic systems is based on handling each measurement signal independently before combining the results at the analysis or decision-making stage. This approach aims to improve monitoring and diagnostic accuracy due to the nonlinear and time-varying nature of signals generated by photovoltaic systems.

Each signal produced by the photovoltaic system is treated as an independent channel, such as voltage, current, temperature, and solar irradiance signals. Each signal undergoes separate preprocessing steps, including noise removal, drift correction, temporal resampling, or normalization.

After that, a set of distinctive features is extracted from each signal individually, such as the mean value, instantaneous variation, variance, and spectral characteristics. These features are used to analyze the individual behavior of each signal in order to detect abnormal deviations or changes related to environmental or operational conditions.

In the final stage, a late fusion approach is applied, where the results of individual signal analysis are combined instead of merging the signals themselves. This method contributes to improving fault detection, performance estimation, and the enhancement of maximum power point tracking (MPPT) algorithms. [25]

2.3.4 Limitation of Conventional Methods

Conventional fault detection methods suffer from several limitations [23]:

- a. Limited ability to detect faults early
- b. Difficulty in distinguishing between types of faults

- c. Low accuracy in large and complex systems
- d. Heavy reliance on manual inspection
- e. Performance affected by environmental conditions
- f. Insufficient data used for diagnosis.

2.4 Model-Based Fault Detection Methods

- **Model-Based Approaches:**

Model-based methods rely on mathematical models of the PV system derived from physical laws and electrical circuit equivalence. These methods compare measured values with model-predicted values; significant deviations indicate faults [17].

2.4.1 Mathematical Modeling of PV Systems

As detailed in Chapter I (Section 1.3.4), the single-diode model is used to represent PV system behavior. The output current is given by:

$$I = I_{ph} - I_0 \left(e^{\frac{q(V+RSI)}{AKT}} - 1 \right) - \frac{V + RSI}{Rsh} \quad \text{II. 1}$$

2.4.2 Residual Generation Techniques

Residual generation techniques are considered one of the most important model-based fault detection methods in photovoltaic systems, as they rely on comparing the actual measured values with the estimated values obtained from a mathematical model representing the healthy state of the system. The residual is mathematically defined as follows:

$$r(t) = y_{(measured)}(t) - y_{(model)}(t) \quad \text{II. 2}$$

Where $y_{measured}$ represents the measured signals (such as voltage, current, or power), and y_{model} represents the signals estimated from the model.

- **In the normal condition:** $r(t) \approx 0$
- **In the case of a fault:** $r(t) \neq 0$

The detection process goes through three main stages: residual generation, evaluation by comparing it to a threshold, and decision making. The fault detection condition can be expressed as follows:

$$\begin{aligned} \text{If } |r(t)| > \delta &\Rightarrow \text{fault exists} \\ \text{If } |r(t)| \leq \delta &\Rightarrow \text{system is healthy} \end{aligned}$$

Where δ represents the threshold defined based on noise and operating conditions.

This technique is characterized by its ability to detect faults early and its applicability in real-time operation. However, its effectiveness depends on the accuracy of the used model and is affected by noise and environmental variations, which sometimes requires the use of improvement techniques such as observers or adaptive thresholds. [26]

2.4.3 Observer-Based Methods

Observer-based methods are a class of model-based fault detection techniques. They construct a mathematical model to estimate outputs or unmeasured internal states. The residual is generated by comparing estimated and measured values [20].

Mathematically, if the system is described as follows:

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) \\ y(t) = h(x(t)) \end{cases} \quad \text{II. 3}$$

Where:

$x(t)$: state vector

$u(t)$: input vector

$y(t)$: output vector

The observer estimates the states according to:

$$\begin{cases} \dot{\hat{x}}(t) = f(\hat{x}(t), u(t)) + L (y(t) - \hat{y}(t)) \\ \hat{y}(t) = h(\hat{x}(t)) \end{cases} \quad \text{II. 4}$$

Where:

$\hat{x}(t)$: estimated states

$\hat{y}(t)$: estimated outputs

L : observer gain matrix

The residual is defined as follows:

$$r(t) = y(t) - \hat{y}(t) \quad \text{II. 5}$$

2.4.4 Advantages and Disadvantages

Model-based methods have the following characteristics [26]:

a. Advantages:

- High accuracy in fault detection
- Ability to detect faults at an early stage
- Capability of fault diagnosis
- Does not rely on large historical datasets

b. Disadvantages:

- Strong dependence on model accuracy
- Difficulty in accurately modeling real systems
- Need for continuous calibration
- Sensitivity to noise and uncertainty
- Mathematical and computational complexity. [23]

2.5 Data-Driven Fault Detection Methods

The data-driven approach relies on analyzing actual operational data without the need for precise mathematical models. This approach uses machine learning and artificial intelligence techniques to discover relationships between variables such as solar irradiance, voltage, and current, to recognize the normal behavior of the system, and then detect any deviation that may indicate the presence of a fault.

This method is characterized by its ability to handle complex systems and operate in real time. It also offers flexibility through the use of supervised and unsupervised learning methods as well as anomaly detection techniques, making it one of the most effective approaches for monitoring modern photovoltaic systems. [27]

2.5.1 Statistical-Based Methods

Statistical techniques analyze operational data (voltage, current, power, temperature, irradiance) to identify deviations from normal system behavior. They are particularly effective for multivariate systems and early fault detection.

Principal Component Analysis (PCA) is one of the most widely used statistical techniques in this field, enabling dimensionality reduction while preserving data variance. Key statistical indicators include:

- Hotelling's T^2 : Monitors variations within the principal component subspace
- Squared Prediction Error (SPE): Monitors residuals outside the PCA model

Additional tools such as CUSUM and EWMA control charts track variable evolution over time to detect slow changes affecting system efficiency.

Statistical techniques are suitable for intelligent monitoring and predictive maintenance due to their simplicity, low computational cost, and real-time implementation capability [28].

2.5.2 Comparison of Fault Detection Techniques

Table 2.1: Theoretical Fault Isolation and Diagnostic Theoretical Matrix for the PV System.

Method	Advantages	Limitations
PCA	Simple, fast, dimensionality reduction	Limited to linear systems
KPCA	Handles nonlinear systems	High computational complexity
SVM	Good classification accuracy	Sensitive to parameter tuning
CNN	High detection accuracy	Requires large datasets
LSTM	Suitable for time-series analysis	Long training time

This thesis focuses on PCA and its dynamic extension (DPCA) due to their simplicity, effectiveness, and suitability for real-time monitoring applications.

2.5.3 Principal Component Analysis (PCA)

One of the most used multivariate statistical methods for fault detection, dimensionality reduction, and data analysis is Principal Component Analysis (PCA). PCA preserves the majority of the significant information in the original dataset while converting the original correlated variables into a new collection of uncorrelated variables known as principle components (PCs).

By projecting the original variables into a lower-dimensional subspace, PCA's primary goal is to simplify huge datasets. By optimizing the variance that the new principal components capture, this transformation is accomplished. The variance is greatest in the first principal component, and the remaining variance is captured in decreasing order by the subsequent components.

Numerous electrical and environmental factors, including voltage, current, power, temperature, and solar irradiation, are strongly connected in photovoltaic systems. Consequently, PCA offers an effective way to find anomalous operating situations and extract the most pertinent data.

- **Mathematical Formulation of PCA:**

Let $\mathbf{X}$ be the normalized data matrix of size $n \times m$, where n is the number of observations and m is the number of variables.

❖ **Step 1: Normalization (Z-score)**

$$X_{norm} = \frac{X - \mu}{\sigma} \quad \text{II. 6}$$

Where:

- μ is the mean vector
- σ is the standard deviation vector

❖ **Step 2: Covariance Matrix**

$$\Sigma = \frac{1}{n-1} X_{norm}^T X_{norm} \quad \text{II. 7}$$

❖ **Step 3: Eigenvalue Decomposition**

$$\Sigma = V \Lambda V^T \quad \text{II. 8}$$

Where:

- V is the matrix of eigenvectors (loadings)
- Λ is the diagonal matrix of eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$

❖ **Step 4: Projection onto Principal Components**

$$T = X_{norm} P \quad \text{II. 9}$$

Where:

- P is the matrix of selected principal components (first k columns of V)
- T is the score matrix

❖ **Step 5: Reconstruction and Residuals**

$$\hat{X} = T P^T = X_{norm} P P^T \quad \text{II. 10}$$

$$E = X_{norm} - \hat{X} = X_{norm} (I - P P^T) \quad \text{II. 11}$$

Where:

- $\hat{X}$ is the reconstructed data
- E is the residual matrix

❖ **PCA Algorithm Steps:**

- Data collection and preparation
- Scaling and normalization (Z-score)
- Covariance matrix calculation
- Eigenvalue and eigenvector extraction
- Principal component selection (using CPV criterion)
- Projection of original data into PCA subspace

- Statistical index calculation for fault detection

PCA is widely used due to its simplicity, low computational cost, and efficiency with multivariate correlated systems. However, it assumes linear correlations between variables and may not adequately describe highly nonlinear systems [9].

2.5.4 Dynamic Principal Component Analysis (DPCA)

Dynamic Principal Component Analysis (DPCA) extends traditional PCA to address one of its main limitations: the inability to capture temporal correlations between observations.

DPCA augments the data matrix by including time-lagged variables. For each observation $x(t)$, the augmented vector includes current and past values.

a. Mathematical Formulation of DPCA:

Let $x(t) \in \mathbb{R}^m$ be the measurement vector at time t . The augmented vector with l time lags is:

$$x_d(t) = [x(t), x(t-1), x(t-2), \dots, x(t-l)] \in \mathbb{R}^{m \times (l+1)} \quad \text{II. 12}$$

The augmented data matrix X_d is constructed as:

$$X_d = \begin{bmatrix} x(t) & x(t-1) & x(t-2) & \dots & x(t-l) \\ x(t-1) & x(t-2) & x(t-3) & \dots & x(t-l-1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x(t+n-1) & x(t+n-2) & x(t+n-3) & \dots & x(t+n-l-1) \end{bmatrix} \quad \text{II. 13}$$

The total number of variables in the augmented matrix is:

$$m_{aug} = m \times (l+1) \quad \text{II. 14}$$

Standard PCA is then applied to the augmented matrix X_d :

$$X_d = T_d P_d^T + E_d \quad \text{II. 15}$$

Where:

- T_d is the dynamic score matrix
- P_d is the dynamic loading matrix
- E_d is the dynamic residual matrix

b. Choice of Time Lag l :

The number of time lags l is typically determined using:

- Autocorrelation function analysis
- Cross-validation

- Process knowledge (in this thesis, $l = 3$ is used)

DPCA captures both spatial and temporal correlations among variables, making it more effective than standard PCA for detecting faults in industrial systems where variables change over time[29].

2.5.5 Statistical Monitoring Indices

a. Hotelling's T^2 Statistic:

Hotelling's T^2 statistic is a statistical indicator used with PCA to monitor variations within the principal component subspace. It quantifies how each observation deviates from normal operating conditions.

❖ Mathematical Definition:

$$T^2 = t^T \Lambda_k^{-1} t \quad \text{II. 16}$$

Where:

- t is the score vector of the observation (dimension $k \times 1$)
- Λ_k is the diagonal matrix of the first k eigenvalues

In matrix form:

$$T^2 = x^T P \Lambda_k^{-1} P^T x \quad \text{II. 17}$$

❖ Control Limit for T^2 :

Under the assumption of multivariate normality, the control limit is given by the F-distribution:

$$T_{thresh}^2 = \frac{k(n-1)}{n-k} F_\alpha(k, n-k) \quad \text{II. 18}$$

Where:

- k is the number of retained principal components
- n is the Sample size
- α is the significance level (typically 0.95, 0.99, or 0.999)
- $F_\alpha(k, n-k)$ is the critical value of the F-distribution

❖ Interpretation:

- Under normal conditions: $T^2 \leq T_{thresh}^2$
- When a fault occurs: $T^2 > T_{thresh}^2$ (indicates abnormal system behavior)

The T^2 statistic is particularly effective for detecting faults affecting the internal PCA model structure, making it valuable for monitoring correlated variables in PV systems [30].

b. Squared Prediction Error (SPE) :

The Squared Prediction Error (SPE), also known as the Q-statistic, measures the residual error between original data and PCA-reconstructed data. It is computed from the residual matrix.

❖ Mathematical Definition:

For a single observation x :

$$\begin{cases} e = x - \hat{x} = (I - PP^T)x \\ SPE = e^T e = \|e\|^2 \end{cases} \quad \text{II. 19}$$

Where:

- e is the residual vector
- $\hat{x}$ is the reconstructed observation
- I is the identity matrix

❖ Control Limit for SPE (Box's Method):

The SPE control limit is approximated using the chi-square distribution:

$$\theta_i = \sum_{j=k+1}^m \lambda_j^i \text{ for } i = 1, 2, 3 \quad \text{II. 20}$$

$$h_0 = 1 - \frac{2\theta_1\theta_3}{3\theta_2^2}$$

$$SPE_{thresh} = \theta_1 \left(\frac{z_\alpha \sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right)^{\frac{1}{h_0}} \quad \text{II. 21}$$

Where:

- λ_j are the eigenvalues of the covariance matrix
- k is the number of retained principal components
- m is the total number of variables
- z_α is the standard normal deviate (e.g., 2.33 for $\alpha = 0.99$)
- h_0 is a parameter adjusting for the chi-square approximation

❖ Simplified SPE Threshold (Approximation):

For practical applications, a simplified threshold can be used:

$$SPE_{thresh} \approx \chi_\alpha^2(h_0) \cdot \frac{\theta_1}{h_0} \quad \text{II. 22}$$

❖ Interpretation:

- Under normal conditions: $SPE \leq SPE_{thresh}$
- When a fault occurs: $SPE > SPE_{thresh}$ (indicates abnormal system behavior)

SPE is more sensitive than T^2 to faults occurring outside the PCA model space, making it particularly useful for detecting sensor malfunctions, unexpected fluctuations, and nonlinear anomalies in PV systems [31].

2.5.6 Fault Localization Techniques

After detecting a fault, localization techniques identify which variables are responsible for the abnormal behavior. This is essential for targeted maintenance and rapid system recovery.

a. Contribution Plot:

A graphical method that determines which variables most significantly contribute to the detected fault. For the SPE statistic, the contribution of variable j is:

$$SPE_j = e_j^2 = (x_j - \hat{x}_j)^2 \quad \text{II. 23}$$

For the T^2 statistic, the contribution of variable j is:

$$T_j^2 = \frac{t_j^2}{\lambda_j} \quad \text{II. 24}$$

Variables with higher contribution values are considered the primary causes of abnormal operation. This technique is useful for detecting sensor malfunctions, voltage disturbances, and current irregularities [31].

b. Reconstruction-Based Contribution RBC:

The RBC method isolates faults by reconstructing the sample vector along the direction of each variable. The contribution of variable j is:

$$RBC_j = \frac{(e^T \xi_j)^2}{\xi_j^T (I - PP^T) \xi_j} \quad \text{II. 25}$$

Where:

- e is the residual vector
- ξ_j is the j -th column of the identity matrix (selection vector)
- P is the loading matrix

RBC is more robust than standard contribution plots as it reduces the "fault smearing effect" where fault energy leaks into correlated variables.

c. Reconstruction Error Map:

A visualization method that displays the distribution of reconstruction errors across variables and samples. Higher reconstruction errors indicate anomalous areas associated with faults. A color map typically depicts error distribution, with high-intensity areas indicating fault locations.

2.6 Conclusion

This chapter has presented a theoretical overview of fault detection methods for photovoltaic systems. Various techniques were covered, including conventional methods (threshold-based, rule-based, signal processing), model-based methods, and data-driven methods. The primary focus was on data-driven statistical techniques due to their efficiency in monitoring multivariate systems and identifying abnormal operating conditions.

Principal Component Analysis PCA and its dynamic extension, Dynamic Principal Component Analysis DPCA, were presented as powerful statistical methods for dimensionality reduction and fault detection in PV systems. The statistical monitoring indices Hotelling's T^2 and Squared Prediction Error SPE were introduced as effective tools for detecting anomalous system behavior.

To enhance fault interpretation and localization, techniques including Contribution Plots and Reconstruction-Based Contribution RBC were introduced.

Based on this theoretical review, the PCA and DPCA methods are selected for implementation in the next chapter, where they will be applied to real operational data from a PV plant to detect and isolate sensor faults.

CHAPTER III

**Implementation of the proposed monitoring
framework**

3.1 Introduction

This chapter presents the implementation, comparative analysis, and validation of data-driven process monitoring frameworks for photovoltaic PV systems. Real meteorological and operational data from the PV Station in Souk Ahras, Algeria, are used to evaluate and compare Conventional Static Principal Component Analysis PCA and Advanced Dynamic PCA (DPCA). The main objective is to build a resilient, automated diagnostic tool capable of isolating sensor abnormalities under fluctuating climatic conditions.

The engineering pipeline systematically covers: data acquisition, Z-score standardization, orthogonal subspace projection via Singular Value Decomposition SVD, and the calculation of statistical control limits (T^2 and SPE). Furthermore, fault identification is achieved using the Reconstruction-Based Contribution RBC method and classification matrices. This practical validation rigorously demonstrates the superiority of the proposed dynamic framework over the static baseline when subjected to a complex, hybrid sensor fault consisting of simultaneous Bias and Limited Drift failures.

3.2 Description of the Oued Kebarit Photovoltaic Station:

3.2.1 Site Description

The Oued Kebarit photovoltaic station is located in the Souk Ahras province, Algeria. This facility operates under fluctuating environmental conditions characterized by significant seasonal variations, elevated ambient temperatures, and moving cloud covers. Such environmental dynamics directly impact the performance and aging behavior of the photovoltaic arrays, making automated fault monitoring and diagnosis crucial for sustaining optimal yield. Consequently, this station serves as an ideal, real-world benchmark for validating data-driven Multivariate Statistical Process Monitoring MSPM techniques under complex weather profiles.

3.2.2 Dataset Description

The experimental validation is conducted using a specialized dataset collected from the operational facility during the month of January. The raw matrix encompasses $n = 816$ synchronized temporal samples (after removing 83 severe outliers from the electrical Power Index) across $m = 6$ critical variables. These variables capture the internal electrical performance alongside ambient meteorological metrics, embedding the essential dynamic cross-correlations required for training the PCA and DPCA frameworks.

The structural configuration and physical units of these monitored variables are detailed in Table 3.1.

Table 3.1: Monitored variables of the Oued Kebarit PV installation. monitored variables.

Variable Index	Parameter Name	Physical Unit	Description
Var 1	Power Index (P)	Kilowatt (kW)	Generated electrical power output
Var 2	Solar Irradiance (G)	Watts per square meter (W/m ²)	In-plane global solar radiation
Var 3	Ambient Temperature (T)	Degrees Celsius (°C)	Surrounding atmospheric temperature
Var 4	Wind Speed (Ws)	Meters per second (m/s)	Speed of wind over the PV arrays
Var 5	Relative Humidity (RH)	Percentage (%)	Moisture percentage in the air
Var 6	Atmospheric Pressure (Patm)	Hectopascal (hPa)	Ambient barometric pressure

3.3 Proposed Fault Detection and Isolation FDI Architecture

To compare static and dynamic monitoring performance, the operational workflow is methodically divided into two concurrent analytical tracks following a structured modular sequence:

a. Data Acquisition and Preprocessing: Continuous multi-channel observations are retrieved from the Souk Ahras facility, followed by structural data cleaning to eliminate outliers, constant null variances, and missing values.

b. Statistical Normalization (Auto-scaling): To prevent high-scale variables from dominating the monitoring space, raw physical attributes are transformed into standardized dimensions with zero-mean and unit-variance.

c. Algorithmic Branching (Static vs. Dynamic):

- **Path A** (Conventional Static PCA): Computes static thresholds and residuals by performing direct Singular Value Decomposition SVD on the instantaneous normalized matrix.
- **Path B** (Advanced Dynamic PCA): Augments the normalized matrix with time-lagged observations Lag = 3 to capture the system's dynamic memory and process transitions before factorization.

d. Statistical Space Monitoring: Concurrent deployment of Hotelling's T^2 control charts for principal subspace variations and Squared Prediction Error SPE charts for residual anomalies.

e. Fault Isolation and Localization: Execution of Reconstruction-Based Contribution (RBC) maps to isolate and rank the exact compromised sensor channels.

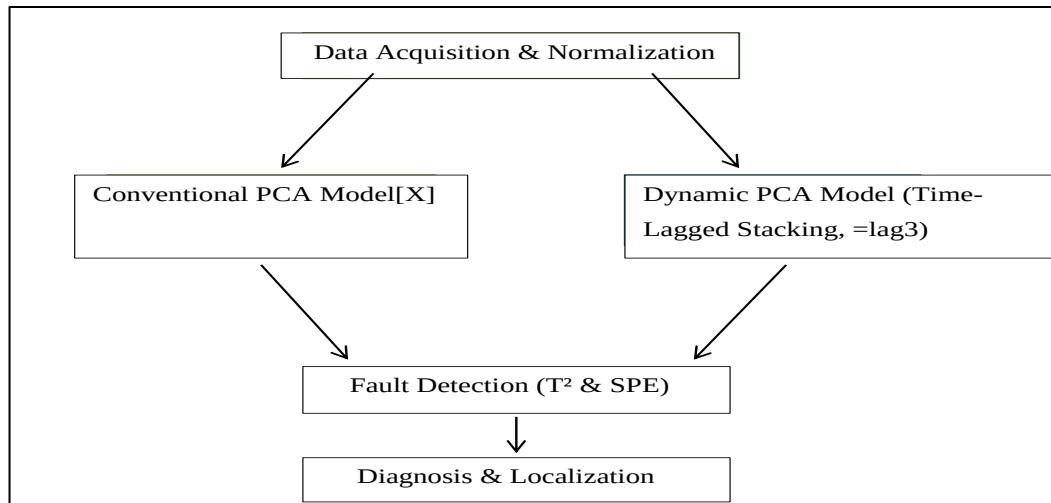


Figure 3.1: Flowchart of the Proposed Static and Dynamic PCA-Based FDI Framework for the PV System.

3.3.1 Data Acquisition Layer

The data acquisition layer represents the first stage of the proposed framework. Its primary function is to collect and organize the operational and environmental measurements required for monitoring. In this study, real-world data were obtained from the Oued Kebarit PV station in Souk Ahras, where six physical parameters are continuously recorded: electrical power output, solar irradiance, ambient temperature, wind speed, relative humidity, and atmospheric pressure.

Since the performance of PV installations is strongly influenced by weather variations, integrating both electrical and meteorological variables allows for an accurate assessment of system behavior. The acquired measurements are subsequently transmitted to the preprocessing

stage. The reliability of the entire monitoring architecture depends significantly on the fidelity of this layer[11].

3.3.2 Data Preprocessing Layer

The data preprocessing layer is essential to guarantee data quality and integrity. Raw measurements from the PV station often include missing values, outliers, or noise induced by sensor degradation and communication anomalies. A preliminary cleaning operation is executed to eliminate variables with zero variance and handle missing observations (outliers cleaning removed 83 samples from the Power Index).

Subsequently, Z-score normalization is applied to standardize all parameters. This step eliminates the disparity in measurement scales, ensuring that each monitored parameter contributes equally to the core of the PCA modeling.

3.3.3 PCA Modeling Layer

The PCA modeling layer reduces the dimensionality of the preprocessed dataset while extracting the most relevant operational information. Principal Component Analysis transforms the original correlated variables into a smaller set of uncorrelated orthogonal components that retain the majority of the system's variability.

By analyzing the covariance structure through eigenvalue decomposition, only the principal components satisfying the cumulative percent variance CPV criterion are retained. This streamlined representation simplifies the monitoring process while capturing the fundamental features of the PV system, serving as the mathematical baseline for constructing the statistical control limits

3.3.4 Fault Detection and Localization Layer

The fault detection and localization layer constitutes the decision-making stage of the architecture. Following the establishment of the statistical baseline, the Squared Prediction Error *SPE* and Hotelling's T^2 statistics are continuously computed and tracked.

These indicators are compared against analytical control thresholds at predetermined confidence levels 95% or 99% to detect anomalous conditions. Once a threshold breach occurs, the Reconstruction-Based Contribution RBC method is triggered to isolate the specific faulty variables, allowing operators to diagnose sensor abnormalities in real time

3.4 Practical Implementation and Subspace Parameterization

Following the theoretical background detailed in Chapter II, this section parameterizes and deploys the statistical monitoring layers specifically tailored for the Oued Kebarit PV installation.

3.4.1 Data Standardization Implementation

The continuous raw variables $m = 6$ exhibit fundamentally different physical dimensions and ranges (e.g., Solar Irradiance reaching 1000 W/m^2 while Ambient Temperature fluctuates below 25°C). To ensure scale-independence, the dataset is normalized using the empirical mean vector μ and standard deviation vector σ established during the nominal training window, according to the Z-score formulation:

$$[Normalized_{Matrix}] = \frac{(Raw_{Matrix} - \mu)}{\sigma} \quad \text{III. 1}$$

3.4.2 Static PCA Subspace Partitioning

Singular Value Decomposition SVD was applied to the standardized static covariance matrix Σ . To determine the optimal number of principal components k , the Cumulative Percent Variance CPV profile was analyzed.

While the first principal component explains 34% of the total variance, the intersection with the strict 95% engineering threshold is achieved by retaining exactly four principal components $k = 4$. These 4 components capture 96.57% of the total systemic variance, successfully isolating the essential cross-correlations between weather conditions and electrical output into the Principal Subspace (P). The remaining analytical variance 3.43% is projected onto the Residual Subspace ($\tilde{P}$) as ambient measurement noise.

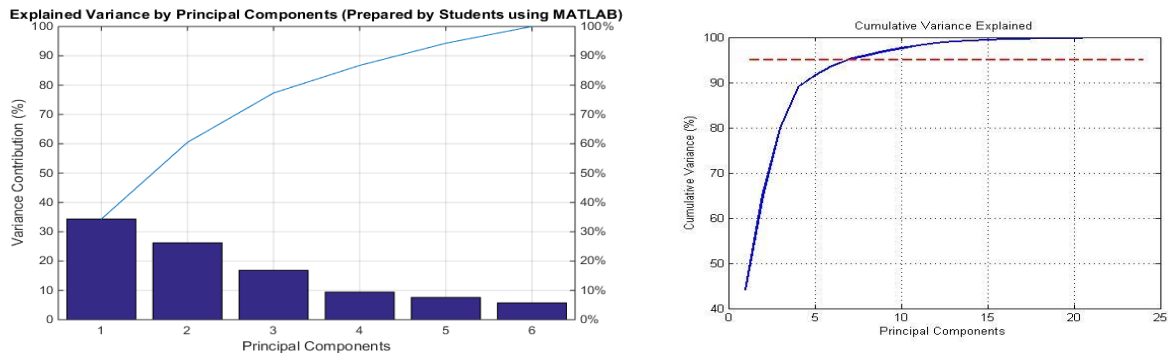


Figure 3.2: Explained variance and cumulative percent variance by each principal component.
(MATLAB simulation results)

3.4.3 Dynamic PCA Spatio-Temporal Expansion

To incorporate the thermodynamic lags and serial auto-correlations characteristic of the desert environment in Souk Ahras, the standardized observation space is extended into the dynamic domain using the time-lagged mathematical model defined in Section 2.5.4.

By applying a time-lag factor (Lag = 3), the instantaneous 6-dimensional observation vector expands into a 24-dimensional spatio-temporal feature space:

$$\{\textit{Total Dynamic Features}\} = 6 \times (3 + 1) = 24$$

Applying SVD to this augmented dynamic matrix shifts the CPV profile. To maintain the same 95% statistical criteria, the dynamic framework retains exactly 7 principal components, which account for 95.10 % of the total dynamic variance, yielding a highly sensitive monitoring baseline.

3.5 Fault Detection Monitoring Limits and Performance Metrics

The streaming testing data is evaluated in real-time by projecting the samples onto the parameterized static and dynamic subspaces to compute the monitoring statistics.

3.5.1 Parameterized Control Chart Thresholds

The statistical control limits for both frameworks are established at a strict 99% confidence level ($\alpha = 0.01$). Based on the theoretical distributions detailed in Section 2.5.5, the baseline thresholds are configured as follows:

- **Static PCA Framework:**

The Hotelling's T^2 control limit is fixed at 13.4187, and the Squared Prediction Error (SPE) threshold is established via Box's approximation at 1.1896.

- **Advanced Dynamic PCA Framework:**

Due to the expanded dynamic subspace dimensionality, the T^2 limit scales to 18.7691, while its corresponding residual SPE threshold stabilizes at 3.0828.

3.5.2 Industrial Evaluation Metrics

To quantitatively evaluate and benchmark the reliability, sensitivity, and speed of the developed control charts before analyzing the experimental curves, three core industrial process performance metrics are formally defined as:

- **False Alarm Rate (FAR):**

Measures the percentage of normal samples incorrectly flagged as anomalies due to noise:

$$FAR = \frac{\text{Number of Normal Samples exceeding Threshold}}{\text{Total Number of Normal Samples}} \times 100\% \quad \text{III. 2}$$

- **Fault Detection Rate (FDR):**

Quantifies the sensitivity of the index by calculating the percentage of actual faulty samples correctly identified:

$$FDR = \frac{\text{Number of Faulty Samples exceeding Threshold}}{\text{Total Number of Faulty Samples}} \times 100\% \quad \text{III. 3}$$

- **Detection Delay (τ):**

Measures the lag in samples between fault injection (t_f) and its statistical detection (t_d):

$$\tau = t_f - t_d \quad \text{III. 4}$$

3.6 Experimental Results and Discussion

3.6.1 Normal Operating Baseline and Information Retainment

Before assessing the frameworks under faulty conditions, the monitoring models are calibrated using the healthy, nominal operational dataset from the Oued Kebarit PV plant. This baseline phase verifies the data-driven models' capacity to capture nominal environmental and electrical dynamics without false status changes.

The original dataset is cleaned of outliers and standardized to eliminate scaling differences between varying physical units. The decomposition of the standardized data matrix yields the following baseline structural configuration:

- Conventional Static PCA:

Retains $k = 4$ principal components out of 6 variables to capture 96.57% of the global variance, discarding the remaining 3.43% as residual noise.

- Advanced Dynamic PCA (DPCA):

By introducing a time lag of $\text{Lag} = 3$, the measurement space expands into a $24\{-dimensional\}$ augmented feature space. The dynamic model retains $k = 7$ principal components to encompass 95.10% of the cumulative process variance.

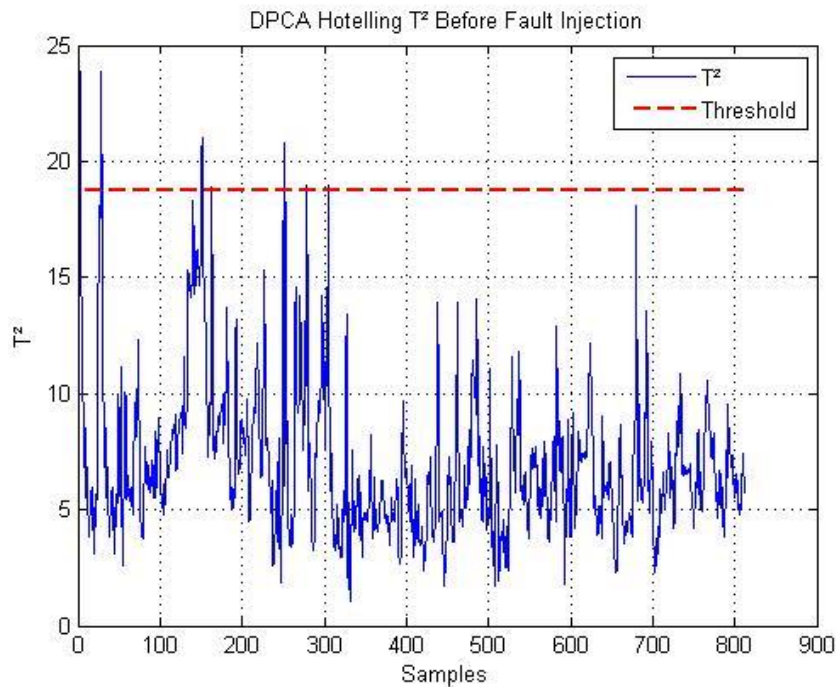


Figure 3.3: DPCA Hotelling T^2 Before Fault Injection (MATLAB simulation results)

To verify the steady-state modeling visually before injecting the anomalies, the standardized multi-channel profiles are tracked across the initial nominal window (samples 1 to 299). All process variables exhibit standard statistical noise fluctuating around zero, confirming a clean baseline and perfect tracking behavior.

3.6.2 Physically-Consistent Fault Injection Scenario

To evaluate the operational resilience of the monitoring frameworks, a multi-sensor failure profile is introduced into the Oued Kebarit PV asset dataset (816 samples, 6 variables). The temporary fault is strictly injected between samples 300 and 600.

Instead of arbitrary noise, a physically-consistent hybrid fault is simulated to model realistic plant thermodynamics where environmental conditions and electrical outputs are coupled:

a. Solar Irradiance Channel (Variable 2):

Subjected to a negative step bias of -2.2 combined with a progressive linear drift of -0.01 per sample. This models sudden cloud shading followed by dust accumulation on the pyranometer.

b. Power Index Channel (Variable 1):

Subjected to a proportional negative step bias of -1.2 and a linear drift of -0.005 per sample, representing the physical drop in electrical generation caused by the irradiance loss.

The linear drift growth is capped at a saturation threshold at sample **drift_end = 450**. Beyond this sample, the fault magnitude remains stable until complete environmental clearance occurs at sample 600. The programmatic setup is structured in Table 3.2.

Table 3.2: Programmatic properties of the injected hybrid fault

Physical Parameter	Channel Vector	Applied Anomaly Type	Numerical Magnitude	Temporal Window
Power Index	Column 1	Step Bias + Bounded Drift	$-1.2 - 0.005 \times t_{drift}$	Samples 300 to 600
Solar Irradiance	Column 2	Step Bias + Bounded Drift	$-2.2 - 0.010 \times t_{drift}$	Samples 300 to 600

3.6.3 Comparative Fault Detection and Statistical Threshold Analysis

This phase evaluates the real-time detection performance of both frameworks during the active fault window (samples 300 to 600). The execution profiles of Hotelling's T^2 and Squared

Prediction Error (SPE) control charts are analyzed to verify threshold breaches and evaluate latency.

a. Conventional Static PCA Monitoring Results:

In the conventional static PCA framework, the monitoring indicators project the process variance onto separate subspaces based on a 99% analytical confidence limit. The control boundaries are calculated as: **T^2 Limit = 13.4187** and **SPE Limit = 1.1896**.

- **T^2 Chart Behavior:** The static T^2 chart responds instantaneously at sample 300 without any latency (**Delay = 0 samples**). However, it fails to maintain the breach, capturing only **41.86%** of the active fault window.
- **SPE Chart Behavior:** The static SPE chart exhibits a severe limitation in tracking the hybrid fault. It fails to detect the initial shock and introduces an operational latency of **29 samples** before crossing the threshold line. It captures only **65.45%** of the total faulty samples.

b. Advanced Dynamic PCA (DPCA) Monitoring Results:

The advanced DPCA framework incorporates time lags (Lag = 3) to account for serial correlations. The calculated statistical limits for this augmented space are: **T^2 Limit = 18.7691** and **SPE Limit = 3.0828**.

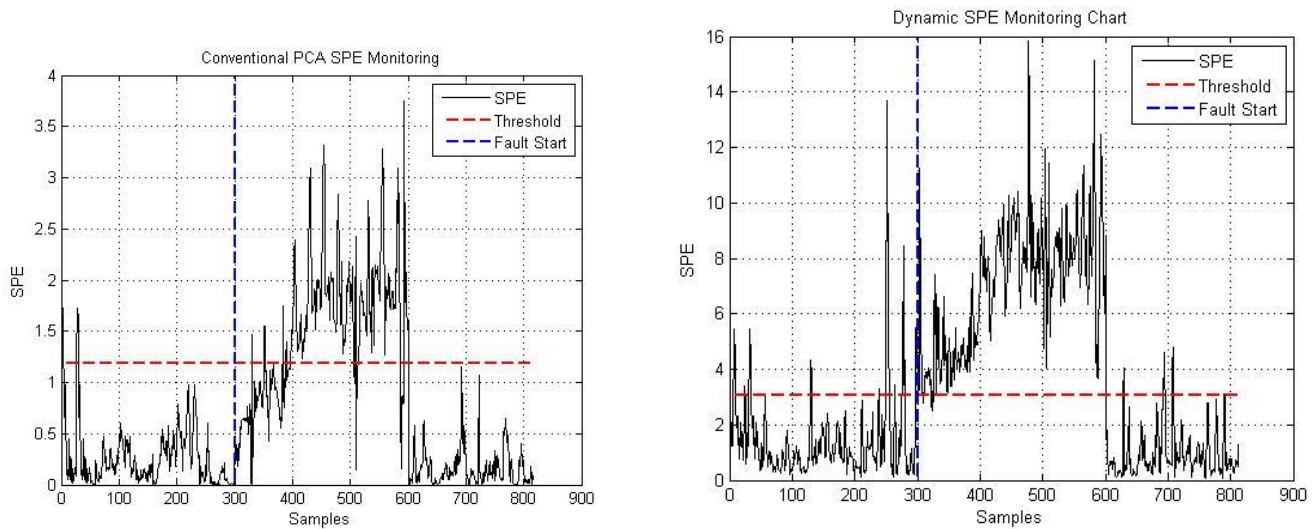


Figure 3.4: Monitoring Charts for the PV System After Fault Injection

T² Chart Behavior: The dynamic T² chart reacts with a minimal delay of **3 samples**, achieving a Fault Detection Rate (FDR) of **55.48%**.

- **SPE Chart Behavior:** The dynamic SPE chart completely eliminates the latency observed in the static baseline, registering a clean, vertical threshold breach at the exact instance of fault injection (**Delay = 0 samples**). Driven by the dynamic residual tracking, it maintains a near-flawless Fault Detection Rate of **98.67%** across the entire window. Furthermore, upon reaching sample 601, the chart immediately resets below the control limit, confirming excellent automation stability without residual memory smearing.

To summarize the operational efficiency before moving to the diagnostic layer, the statistical parameters computed from both validation routines are compiled and contrasted in Table 3.3.

Table 3.3: Quantitative statistical fault detection and latency metrics comparison

Monitoring Framework	Control Chart Index	Analytical Limit Threshold	False Alarm Rate (FAR %)	Fault Detection Rate (FDR %)	Empirical Detection Delay (τ)
Conventional Static PCA	T ² Chart	13.4187	2.01%	41.86%	0 Samples
	SPE Chart	1.1896	2.68%	65.45%	29 Samples
Advanced Dynamic PCA	T ² Chart	18.7691	1.76%	55.48%	3 Samples
	SPE Chart	3.0828	7.03%	98.67%	0 Samples

The quantitative results in Table 3.3 show that the conventional static SPE chart introduces a dangerous blind spot of 29 samples due to its inability to decouple temporal dynamics from unmodeled residuals. The advanced DPCA framework overcomes this limitation, achieving immediate detection ($\tau = 0$) and a high detection rate (98.67%). This confirms the efficacy of expanding the observation space with time-lagged variables to detect physically-consistent hybrid faults.

3.6.4 Diagnostic Fault Localization and Variable Isolation

Once a statistical threshold breach is confirmed by the detection layer, the diagnostic engine automatically computes the reconstruction-based contribution profiles to isolate the specific variables responsible for the plant disturbance.

a. Conventional Static PCA Fault Localization:

In the conventional static model, the localization index is calculated by analyzing the mean squared reconstruction errors (E_{fault}) strictly within the active fault window (samples 300 to 600). The physical parameters are ranked based on their absolute empirical contribution values.

- **Empirical Analysis of Static Ranking:** The static model successfully identifies the affected channels; the Irradiance variable ranks first with a contribution value of **0.700531**, followed immediately by the Power Index in the second position with **0.679113**. In contrast, the remaining nominal environmental channels (Temperature, Humidity, Pressure, and Wind Speed) remain isolated near the zero baseline, proving accurate decoupling of the background process noise.

b. Advanced Dynamic PCA (DPCA) Fault Localization:

The advanced DPCA framework aggregates the squared reconstruction error contributions (SREC) across the augmented 24-dimensional feature space. By summing the dynamic sub-components ($cols = v:m_0:m$) associated with each original physical channel, the global dynamic localization profile is extracted.

- **Empirical Analysis of Dynamic Ranking:** The mathematical superiority of DPCA is demonstrated by expanding the quantitative gap between faulty and healthy channels; the Irradiance contribution index increases to **2.985119** (Rank 1), and the Power Index reaches **2.404929** (Rank 2). This numerical amplification enhances diagnostic decision-making reliability within commercial PV control rooms by preventing ambiguous root isolation

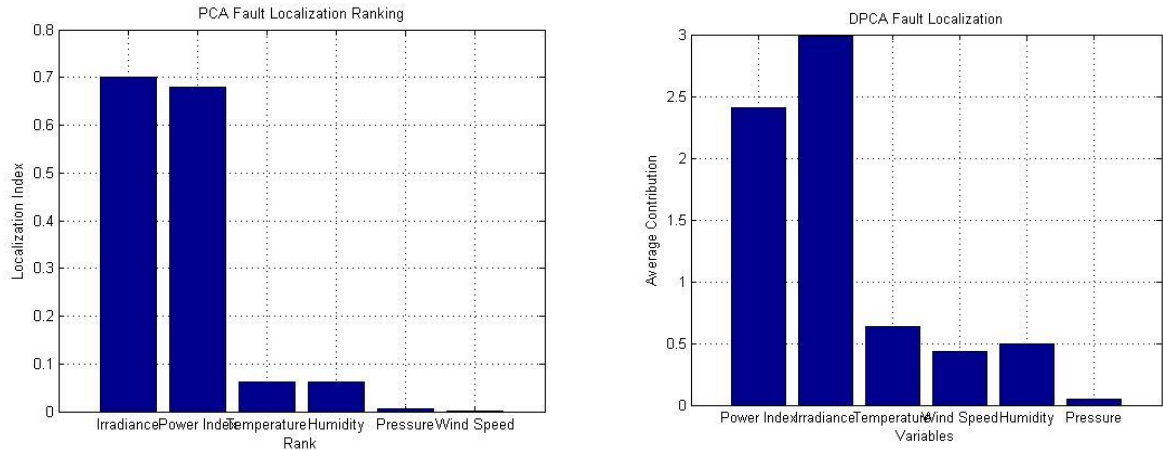


Figure 3.5: Comparative Diagnostic Variable Ranking Matrix (MATLAB simulation results)

To benchmark the diagnostic layers, the numerical localization indices and the subsequent physical component rankings for both frameworks are consolidated in Table 3.4.

Table 3.4: Multi-variable fault localization index and diagnostic ranking comparison

Variable Rank	Physical Parameter Channel	Conventional PCA Index	CPCA Status	Advanced Dynamic PCA Index	DPCA Status
1	Solar Irradiance	0.700531	Primary Root Fault	2.985119	Primary Root Fault
2	Power Index	0.679113	Secondary Induced Fault	2.404929	Secondary Induced Fault
3	Temperature	0.063097	Unaffected Baseline	0.635303	Unaffected Baseline
4	Humidity	0.062283	Unaffected Baseline	0.495546	Unaffected Baseline
5	Wind Speed	0.001131	Unaffected Baseline	0.432033	Unaffected Baseline
6	Pressure	0.005206	Unaffected Baseline	0.049803	Unaffected Baseline

Spatio-Temporal Trajectory Analysis of Top 10 Dynamic Features

To reveal the internal mechanics of the dynamic lag window (Lag = 3), the individual contribution scores of the top 10 most influential dynamic features within the 24-dimensional space are extracted and analyzed from the DPCA framework:

- a. Feature 2 (Irradiance - Instantaneous t): Contribution = 1.3626
- b. Feature 13 (Power Index - Lag 2): Contribution = 0.9852
- c. Feature 19 (Power Index - Lag 3): Contribution = 0.9111
- d. Feature 8 (Irradiance - Lag 1): Contribution = 0.7540
- e. Feature 20 (Irradiance - Lag 3): Contribution = 0.4933
- f. Feature 7 (Power Index - Lag 1): Contribution = 0.4203
- g. Feature 14 (Irradiance - Lag 2): Contribution = 0.3753
- h. Feature 23 (Temperature - Lag 3): Contribution = 0.2756
- i. Feature 3 (Temperature - Instantaneous t): Contribution = 0.2755
- j. Feature 21 (Temperature - Lag 1): Contribution = 0.2713

The extracted feature distribution explicitly proves that the dynamic framework tracks the historical trajectory of the system rather than inspecting isolated, discrete data points. The dominance of Feature 2 highlights the instantaneous impact of the hybrid shock on the solar irradiance channel at time t . This is immediately backed by Feature 13 and Feature 19, which represent the time-lagged profiles (Lag 2 and Lag 3) of the Power Index channel. Capturing these serial cross-correlations across successive time lags explains why the advanced DPCA framework completely eliminates detection latency ($\tau = 0$). By embedding the thermal and electrical settling memories of the PV arrays directly into the augmented mathematical subspace, the model tracks the continuous degradation path without delayed responses.

3.6.5 Global Classification Performance and Benchmarking Summary

To establish the conclusive engineering verification of this study, a global classification benchmark is executed. This phase evaluates the overall data-driven decision performance by

assessing how effectively each framework discriminates between true healthy states and active multi-variable anomalies across the complete time-series horizon.

The quantitative validation metrics are extracted directly from the empirical confusion matrices of both routines. While the statistical detection phase focused strictly on sub-space threshold behaviors, this global assessment considers the combined mathematical decisions $\left(\mathcal{Y}_{\{pred\}} = (T^2 > T_{\{limit\}}^2) \mid (SPE > SPE_{\{limit\}}) \right)$ to measure the system-wide classification accuracy.

The definitive numerical indicators for both configurations are consolidated and contrasted in Table 3.5.

Table 3.5: Definitive global classification and performance matrix benchmarking

Evaluation Metric Vector	Conventional Static PCA Baseline	Advanced Dynamic PCA (DPCA) Framework	Quantified Performance Optimization (Delta Δ)
Global Classification Accuracy	0.8689	0.9446	+ 7.57% (Accuracy Jump)
Precision Score	0.9369	0.8787	- 5.82%
Recall (Sensitivity) Score	0.6910	0.9867	+ 29.57% (Sensitivity Surge)
Global F1-Score Balance	0.7954	0.9296	+ 13.42% (Total Optimization)

Analytical Interpretation of the Performance Jump

The statistical benchmarking in Table 3.5 highlights a substantial performance increase when transitioning from static modeling to the advanced dynamic framework:

c. The Global Accuracy Jump:

The overall classification accuracy exhibits a clear increase, rising from **0.8689** in the conventional PCA model to **0.9446** in the dynamic PCA model. This **7.57% net improvement** demonstrates the robustness of integrating a time-lagged window (Lag = 3) to capture serially correlated solar asset vectors.

d. Resolving the Recall vs. Precision Trade-off:

In industrial fault detection applications, a critical trade-off often exists between Precision and Recall.

- The conventional static model achieves a high Precision score of **0.9369** but suffers from a low Recall rate of **0.6910**. This indicates a high number of missed detections (False Negatives), primarily caused by the **29-sample delay** of the static SPE chart in detecting the progressive linear drift component.

- The advanced DPCA framework resolves this limitation by increasing the Recall score to **0.9867**—representing a **29.57% sensitivity surge**. By expanding the observation space into a 24-dimensional matrix, the dynamic residual space eliminates the blind spot associated with creeping drift profiles. This ensures that **98.67%** of the active failure window is successfully identified and reported to plant operators.

- F1-Score Optimization:

The global F1-Score, which represents the harmonic mean of Precision and Recall, summarizes the overall detection capability. The F1-Score increases from **0.7954** to **0.9296** (**+13.42% overall optimization**). This improvement confirms that the advanced DPCA framework achieves a superior balance between high sensitivity and low false-alarm distributions.

In conclusion, this empirical evaluation confirms that the proposed Advanced Dynamic PCA framework successfully overcomes the structural limitations of static models. By capturing temporal cross-correlations, it achieves rapid, zero-delay detection and high diagnostic isolation accuracy for physically-consistent hybrid faults in grid-connected solar PV systems.

3.7 Conclusion:

This chapter presented a comparative benchmarking between conventional and advanced dynamic data-driven frameworks using operational solar data to detect physically-consistent failures. The quantitative validation proved that the Conventional Static PCA framework is limited by serial correlations, as its SPE chart introduced a 29-sample detection delay and captured only 65.45% of the fault window, locking the baseline F1-score at 0.7954. Conversely, the Advanced Dynamic PCA (DPCA) framework successfully eliminated this latency by mapping process variables into a 24-dimensional time-lagged feature space ($\text{Lag} = 3$), which achieved immediate detection (0 samples delay), maximized the fault detection rate to 98.67%, and optimized the global classification accuracy to 0.9446 and the F1-score to 0.9296. Furthermore, the automated diagnostic layer precisely isolated Solar Irradiance and Power Index as the core failure root sources, verifying that the advanced DPCA architecture embeds the spatiotemporal memory of the plant directly into its mathematical subspaces to provide superior reliability for real-time industrial PV SCADA applications.

General

CONCLUSION

GENERAL CONCLUSION

With an emphasis on improving the efficiency and dependability of photovoltaic (PV) systems by early fault identification and isolation, this work has offered an advanced practical and evaluative study in the field of data-driven Multivariate Statistical Process Control (MSPC). The main goal of the study was to use the MATLAB environment to perform a thorough quantitative and qualitative comparison between the traditional static Principal Component Analysis (PCA) technique and its dynamic extension, Dynamic Principal Component Analysis (DPCA). Real operating data from the Oued El Kebrit solar power facility in Algeria's Souk Ahras Province was used to assess the suggested approaches.

The analysis of statistical performance indicators and the simulation results led to the following important conclusions:

- The DPCA method can account for temporal dynamics and process operating circumstances, it clearly outperformed the traditional PCA technique in terms of technical performance. This feature improved the Fault Detection Rate (FDR) while drastically lowering the False Alarm Rate (FAR).
- In addition to lowering the Detection Delay, the statistical control charts based on Hotelling's T^2 and Squared Prediction Error (SPE) were successful in identifying and monitoring even small process deviations and faults, such as partial shading conditions and short-circuit faults, as soon as they occurred.
- In the Oued El Kebrit solar power plant, Reconstruction-Based Contribution (RBC) charts shown high accuracy and efficacy in fault isolation and in identifying the variables causing anomalous system behaviour, consequently enabling targeted maintenance activities.

This study concludes that adopting dynamic statistical process monitoring approaches and intelligent data-driven monitoring techniques is a strategic solution for improving the protection and dependability of photovoltaic power plants operating in harsh desert environments. The results show how dynamic monitoring techniques can significantly improve fault identification and diagnostic performance while lowering operational and financial losses.

Building on the successes of this work, a number of intriguing avenues for further research are suggested. In order to handle the intrinsic nonlinearities and complexity of solar systems, they include the application of sophisticated nonlinear dynamic monitoring techniques like Kernel Dynamic Principal Component Analysis (KDPCA). Additionally, the incorporation of deep learning algorithms may improve the accuracy of fault classification and allow for more sophisticated diagnostic capabilities. In order to enable prompt defect identification and quicker autonomous reactions in field-deployed solar power plants, future research should concentrate on integrating these monitoring algorithms into real-time monitoring platforms and industrial supervisory systems.

References

References

- [1] R. A. Messenger and J. Ventre, *Photovoltaic Systems Engineering*, CRC Press, 2010.
- [2] A. Kateregga, "Solar Energy Lecture Notes: Introduction to Photovoltaic Systems," *Scribd*, 2023.
- [3] BENY New Energy, "Professional DC & AC Components and EV Charging Solutions," *BENY Technical Documentation*, 2026.
- [4] Z. Mirza and N. Gupta, "Integration of solar power to the electric grid: A case study," *International Journal of Engineering, Science and Technology*, vol. 14, no. 3, pp. 74–84, 2022.
- [5] E. Halassa, "Contribution to control and monitoring of PV system based on deep learning applied to small-scale solar farms," Doctoral thesis, Dept. Elect. Eng., Ziane Achour Univ. of Djelfa, Algeria, 2024.
- [6] F. Ramdani and N. Hourri, "Study and sizing of photovoltaic systems," Master's thesis, Dept. Physics, Kasdi Merbah University Ouargla, Algeria, 2021.
- [7] PowMr, "Types of solar inverters and which one suits you best," *PowMr Technical Blog*, 2024.
- [8] Chint Global, "Applications of Power Transformers in Solar Energy Systems," *Chint Technical Report*, 2024.
- [9] B. Taghezouit, F. Harrou, Y. Sun, and W. Merrouche, "Model-based fault detection in photovoltaic systems: A comprehensive review and avenues for enhancement," *Renewable and Sustainable Energy Reviews*, vol. 170, p. 112960, 2023.
- [10] "Wind Effects on Cooling and Performance of Photovoltaic Arrays: A Case Study Using Computational Fluid Dynamics Modeling," *Energy*, vol. 293, p. 130596, 2024.
- [11] Haina Solar, "Impact of dust on photovoltaic power generation," *Haina Solar Technical Papers*, 2024.
- [12] PVB, "Advanced Solar Energy Optimization Solutions," *PVB White Paper*, 2024.
- [13] H. Shareef, M. M. Islam, and A. Mohamed, "Effect of Prediction Error of Machine Learning Schemes on Photovoltaic Power Trading Based on Energy Storage Systems," *Energies*, vol. 12, no. 7, p. 1249, 2019.
- [14] IEA-PVPS, "Trends in Photovoltaic Applications 2020: Survey Report of Selected IEA Countries between 1992 and 2019," *Report IEA-PVPS*, T1-38:2020, 2020.
- [15] S. R. Madeti and S. N. Singh, "Online fault detection and classification for photovoltaic systems based on Taguchi loss function," *Energy*, vol. 164, pp. 686–703, 2018.

- [16] Y. Zhao, B. Lehman, R. Ball, and J. F. de Palma, "Outlier detection rules for fault detection in solar photovoltaic arrays," in *Proc. IEEE Photovoltaic Specialists Conference (PVSC)*, 2013, pp. 2913–2918.
- [17] A. Mellit, S. Kalogirou, and L. Paridi, "Artificial intelligence techniques for photovoltaic systems: A review," *Soft Computing*, vol. 27, no. 15, pp. 9041–9060, 2023.
- [18] K. T. Chau, "Fault Detection and Diagnosis," in *Electric Vehicle Machines and Drives: Design, Analysis and Application*, IEEE Press & John Wiley & Sons, 2015.
- [19] A. El Amrani, R. El Bachtiri, and M. Tabaa, "A critical review of PV systems' faults with the relevant detection methods," *Nexus, Elsevier*, 2023.
- [20] M. K. Behera, I. Saidi, M. G. Villalva, and T. Khatib, "Recent techniques for photovoltaic system fault detection and diagnosis: A review," *Advances in Applied Energy*, vol. 9, p. 100118, 2023.
- [21] M. Chine, A. Mellit, V. Lughi, and A. Massi Pavan, "A fault detection method for photovoltaic systems based on artificial neural networks," *Renewable Energy*, vol. 66, pp. 121–136, 2014.
- [22] B. Ando, S. Baglio, A. Pistorio, G. L'Episcopo, and N. Savalli, "A low-cost multi-sensor strategy for monitoring photovoltaic systems," *Solar Energy*, vol. 128, pp. 147–155, 2016.
- [23] H. Braun et al., *Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization*, Springer, Cham, Switzerland, 2023.
- [24] K. Ishaque, Z. Salam, and H. Taheri, "Accurate and fast sun tracking system using simple maximum power point tracking," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 10, pp. 4636–4645, 2011.
- [25] B. Aziz, B. Cetiner, J. S. Stein, and M. G. Villalva, "Model-based fault detection in photovoltaic systems: A comprehensive review and avenues for enhancement," *Renewable and Sustainable Energy Reviews*, vol. 170, p. 112960, 2023.
- [26] S. Lu, S. Lu, and Y. Wang, "Deep learning-based fault detection and diagnosis in photovoltaic systems: a detailed review and future directions," *Neural Computing and Applications*, vol. 35, no. 28, pp. 20455–20478, 2023.
- [27] E. Garoudja, F. Harrou, Y. Sun, K. Kara, A. Chouder, and S. Silvestre, "Statistical fault detection in photovoltaic systems," *Solar Energy*, vol. 150, pp. 485–499, 2017.
- [28] M. G. Naik, P. K. Dash, and R. Bisoi, "Fault Detection and Classification in Photovoltaic System Using Data-Driven Techniques: A Review," *Renewable and Sustainable Energy Reviews*, vol. 148, p. 111320, 2021.

References

- [29] H. Wang, P. Li, and M. Zhang, "Data-Driven Fault Detection for Photovoltaic Systems Using PCA-Based Monitoring Techniques," *Energy Reports*, vol. 7, pp. 589–598, 2021.
- [30] Y. Zhang, J. Ma, and X. Chen, "PCA-Based Statistical Monitoring and Fault Detection for Photovoltaic Systems," *Solar Energy*, vol. 224, pp. 548–558, 2021.
- [31] A. K. Samantaray, S. K. Dash, and B. K. Panigrahi, "Fault Detection and Localization in Photovoltaic Systems Using PCA-Based Statistical Monitoring," *Renewable Energy Focus*, vol. 40, pp. 112–121, 2022.