MODEL DEVELOPMENT AND VALIDATION OF A LARGE-SCALE PHOTOVOLTAIC PLANT WITH A DUAL-AXIS TRACKING SYSTEM: CASE OF THE FREE STATE, SOUTH AFRICA

By

MUDIANDAMBU DIDIER KAMBOLO

Dissertation submitted in fulfilment of the requirements for the degree

Master of Engineering in Electrical Engineering

In the Department of Electrical, Electronic and Computer Engineering,

Faculty of Engineering, Built Environment and Information Technology,

Central University of Technology, Free State

Supervisor: Dr P.A. Hohne

Co-Supervisor: Prof. K. Kusakana

DECLARATION

I, Mudiandambu Didier Kambolo, student number ________, hereby declare that this research project, which is submitted by me to the Central University of Technology, Free State, for the degree Master of Engineering in Electrical Engineering, is my individual work and complies with the Code of Academic Integrity, and other policies, procedures, rules, and regulations of the Central University of Technology, Free State. This project has not been submitted by any other individual for fulfilment (or partial fulfilment) of the requirements for the award of any qualification.

MD KAMBOLO Date: 10 December 2025

DEDICATION

This dissertation is dedicated wholeheartedly to my loving children, Badimuwanda Rooter Kambolo and Lusambu Arianna Kambolo, whose forbearance and comprehension have been a motivating force for me throughout this scholarly endeavour. Those long and difficult hours away from you, chasing this dream, were not always easy; yet your affection and smiles fortified me to persevere. This milestone has taught me that endurance and faith can conquer any obstacle.

To my dear wife, Nyamabu Annick Kambolo: your love, encouragement, and sacrifices have been prominent in this success. You kept me strong through the late nights, moments of doubt, and busy schedules. Your confidence in me, even when I lacked it, kept me going. I am eternally grateful for your support, for participating in this journey, and for your infinite patience.

To my dearest mom, whose faith in me brought me through the darkest and hardest times, your prayers, words of encouragement, and sacrifices for me will always be imprinted in my heart. Thank you for the way that you stood by me and for always being there for me, even in times when I almost gave up.

I also dedicate this study to my brothers and sisters, whose silent support, comprehension, and encouragement have been the pillar on which I leaned more times than words can express. Your support and faith in me provided the optimism that propelled me further, even when the way appeared long and not very clear.

Most of all, I offer thanks to my Heavenly Father for guiding me through divine wisdom, patience, and grace in this journey. This dissertation bears the fruits of His faithfulness in me. Without His help and support every step of the way, I would not have achieved this milestone.

ACKNOWLEDGEMENTS

My sincere gratitude extends to the Central University of Technology, Free State (CUT). The CUT has been very generous in supporting my academic pursuits, especially through financial support and assistance.

The fulfilment of this work was feasible only because many individuals worked together on it, whom I wish to thank from the bottom of my heart.

To my supervisor, Dr P.A. Hohne, I would like to express my deepest gratitude for his faith in the quality of the work and for the encouragement demonstrated throughout this long and difficult journey. His support was highly significant in maintaining my dedication to this research.

To Prof. K. Kusakana, my co-supervisor, I would like to extend much appreciation for his valuable guidance, encouragement, and technical expertise throughout this work.

Finally, I would also like to thank from the bottom of my heart my wife, Nyamabu Annick Kambolo, my son, Badimuwanda Rooter Kambolo, my daughter, Lusambu Arianna Kambolo, my mother, Mrs Lusambu Henriette Mangungu, and my brothers and sisters for their long support and encouragement.

Thanks also to each of you for being part in this success. It would not have been completed without all of your aid, tolerance, and cooperation.

ABSTRACT

The production of outdoor photovoltaic (PV) modules is subject to a broad range of parameters that directly influence their energy output. These parameters can generally be categorised into two groups: module-specific features and external environmental conditions. The former pertains to the physical characteristics of the solar modules themselves, including their efficiency, material composition, and the specific technology employed. In contrast, external parameters encompass environmental factors such as wind speed, solar irradiance, ambient temperatures, and geographical location, which may vary significantly between different locations. For instance, elevated temperatures can diminish the efficiency of solar cells, while low wind speeds may result in inadequate cooling of the modules. Additionally, factors such as solar angles, cloud cover, and dirt accumulation (soiling) further complicate the task of accurately predicting the output of PV systems.

The majority of contemporary research on PV systems has focused on static systems, wherein solar modules are fixed at a predetermined tilt and orientation, and models have been developed to forecast energy generation based on a static set of parameters. However, these models frequently fail to accurately represent the realities of dual-axis tracking systems, which possess higher energy output potential by adjusting the angle of panels to follow the sun's horizontal and vertical movement. Although some studies have investigated dual-axis tracking systems, there is a relative paucity of research conducted within the South African context, particularly regarding the unique challenges posed by the country's climate, which includes semi-arid conditions and significant diurnal temperature variations.

The primary objective of this study was to address the research gap by developing and validating an extremely accurate and dependable predictive model for the energy output and performance of dual-axis PV tracking systems in South Africa, with specific application to the CUT installation in Bloemfontein. This research utilised real-world data gathered from a dual-axis PV tracking power plant at the Central University of Technology, Free State (CUT) in the Free State province. An average tracking system was employed, and the data were interpolated to provide a global aggregate for the entire plant. By incorporating data collected from meteorological stations and the Sunny Portal monitoring

system platform, this research offers a comprehensive approach to modelling PV system performance under South African environmental conditions.

A notable aspect of this study was the utilisation of machine learning (ML) techniques to enhance the accuracy and reliability of the predictive model. Various popular ML methods, including linear regression, decision trees, support vector machines, and artificial neural networks, have been explored using MATLAB's Regression Learner App. These methods were evaluated based on their capacity to predict the power output of the PV component from the tracking system, considering environmental parameters such as global horizontal irradiance, direct normal irradiance, diffuse horizontal irradiance, wind speed, temperature, and relative humidity. Among the models assessed, the ensemble tree model was found to be the most effective, as it achieved a coefficient of determination (R^2) of 0.99338 and a root mean square error of 0.39487. This model successfully characterised the complex, non-linear interactions between environmental parameters, thus demonstrating its potential for accurately forecasting the output of the PV component under various operational conditions and time periods.

The newly developed predictive model not only provides exceptional accuracy but also generalises well across a range of conditions, which suggests its applicability for year-round operation. A sensitivity analysis performed on the model highlighted the significant impact of wind speed in reducing the temperature of the PV cells, which enhanced conversion efficiency. This factor has often been overlooked in other research but was identified through this modelling process as a crucial element that contributes to real-world output variability in South Africa's semi-arid climate. The model also effectively accounted for the dynamic behaviour of solar angles, which change throughout the day and year, thus influencing the system's total energy output. These findings further underscore the necessity of incorporating local environmental conditions into PV performance models for more reliable predictions.

In addition to performance prediction, the study addressed the degradation of PV systems over time, which is an essential consideration for long-term system planning and maintenance. An analysis of data spanning from 2019 to 2023 revealed a gradual decline in the annual energy output of the CUT's PV plant, from 41 689.72 kWh in 2019 to 40 038.13 kWh in 2023, which resulted in a performance loss of approximately 3.96%. This degradation was attributed to various factors, including environmental soiling, component ageing, and maintenance delays. The degradation analysis emphasises the importance of ongoing monitoring and maintenance to mitigate the impact of these factors on long-term system efficiency. The predictive model's ability to reflect real-world changes in system performance adds significant practical value for both energy planners and maintenance teams, as it enables them to anticipate performance issues and to proactively implement corrective measures.

Moreover, this research conducted a comprehensive cost analysis of the CUT's PV system to demonstrate whether it constitutes a sustainable and financially viable investment. The system accrued substantial savings during its first five years of operation, amounting to approximately R714 000. The investment has proven to be highly rewarding, with a return on investment of 2.22. This financial analysis, coupled with the environmental benefits of reduced carbon dioxide emissions and diminished reliance on fossil fuels, further strengthens the case for increased investment in solar energy systems at the CUT and other South African universities.

This research makes a significant contribution to the expanding body of literature on South African PV system performance modelling, specifically for dual-axis tracking systems. The validated performance prediction model represents a meaningful advancement towards facilitating decision making in the planning, installation, and operational management of solar PV systems in South Africa. Given South Africa's increasing focus on the adoption of renewable energy sources, particularly through initiatives such as the Renewable Energy Independent Power Producer Procurement Programme, the results and methodologies presented in this study can inform future site selection, investment appraisal, and energy yield forecasting for PV installations.

In addition to its intellectual contributions, this research provides valuable methodologies for optimising solar power in South Africa. Future investigations may extend the model to other provinces with varying climatic conditions and incorporate economic metrics for a more comprehensive long-term perspective on the benefits of PV systems. Additionally, the development of real-time system performance dashboards for optimisation could further enhance the model's forecasting capabilities. The research also advocates for the implementation of soiling and shading models to improve accuracy in areas with high levels of dust accumulation, as these factors can significantly reduce the efficiency of PV systems.

Overall, this study presents a validated and robust tool for assessing and optimising the performance of dual-axis PV trackers in South Africa, thereby contributing to both the body of knowledge through research and to practical energy planning. The findings of this research hold significance for the advancement of renewable energy technologies and the transition towards a cleaner, more reliable energy grid in South Africa and beyond. By integrating advanced modelling techniques, real data, and sustainability considerations, this study establishes a solid foundation for the future optimisation and forecasting of solar energy, thereby promoting a more sustainable energy future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial intelligence
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
CAMS	Copernicus Atmosphere Monitoring Service
CNN	Convolutional neural network
CO_2	Carbon dioxide
CUT	Central University of Technology, Free State
DHI	Diffuse horizontal irradiance
DNI	Direct normal irradiance
GBM	Gradient boosting machine
GHI	Global horizontal irradiance
GWh	Gigawatt-hour
HIMVO-SVM	Hybrid improved multi-variable optimisation support vector machine
IoT	Internet of Things
IRP	Integrated Resource Plan
kg	Kilogram
kW	Kilowatt
kWh	Kilowatt-hour
kWp	Kilowatt-peak
LCOE	Levelised cost of electricity
LSTM	Long short-term memory
m^2	Square metre
MAE	Mean absolute error
MAPE	Mean absolute percentage error
ML	Machine learning
MLR	Multiple linear regression
MW	Megawatt
NASA-SSE	National Aeronautics and Space Administration's Surface Meteorology and Solar Energy
NREL	National Renewable Energy Laboratory

PIML	Physics-informed machine learning
PV	Photovoltaic
PVsyst	Photovoltaic System Simulation Software
R	South African rand
R^2	Coefficient of determination
REIPPPP	Renewable Energy Independent Power Producer Procurement Programme
RETScreen	Renewable Energy Project Analysis Software
RF	Random forest
RF-ANN	Random forest – artificial neural network
RMSE	Root mean square error
ROI	Return on investment
SAM	System Advisor Model
SAURAN	Southern African Universities Radiometric Network
SCADA	Supervisory Control and Data Acquisition
SDG	Sustainable Development Goal
SDM	Single-diode model
SVM	Support vector machine
SVR	Support vector regression
TDM	Two-diode model
USA	United States of America
UV	Ultraviolet

CHAPTER 1:
INTRODUCTION

1.1 BACKGROUND

Photovoltaic (PV) technology has gained significant momentum as a primary solution for addressing the global energy crisis and facilitating a transition towards a more sustainable energy system. The efficiency of PV systems, which convert sunlight into electrical power, plays a crucial role in their adoption and integration into energy grids worldwide. A substantial body of literature has emerged, focusing on calculating and enhancing the efficiency of PV modules under various conditions. Accurate predictions of these systems' performance are essential, particularly when planning large-scale installations for power generation. The efficiency of PV systems is influenced not only by the design and quality of the modules but also heavily by environmental factors, which vary from one geographic location to another and over time.

PV systems, whether fixed or tracker-mounted, are subject to a range of determining factors. The inherent characteristics of the module – specifically, the material (most commonly silicon), the module design, and the efficiency of the solar cells – provide a foundational basis for the module's performance. However, more significant determinants, such as solar irradiance, atmospheric temperature, and wind speed, primarily govern energy output; yet these factors are often overlooked in idealised calculations.

Solar irradiance, defined as the power from the sun received per unit area, is arguably the most critical factor that influences the performance of a PV system. The energy output of a PV system is directly proportional to the amount of solar radiation incident upon the surface of the panels. Solar irradiance is highly location-specific and varies with the time of day, time of year, and atmospheric conditions. Factors such as altitude, latitude, cloud cover, and atmospheric dust significantly affect solar irradiance, which necessitates accurate modelling to predict the performance of PV systems under real-world conditions [1]. In regions characterised by high dust concentrations or recurrent cloud coverage, for instance, the efficiency of a PV system can be significantly impacted.

Furthermore, the angle at which solar radiation strikes the PV module also influences energy conversion efficiency. The optimal angle of incidence may vary with seasonal changes and the orientation of the solar array, which can be fixed or dynamically adjusted through tracking systems. In fixed systems, the tilt and azimuth angle remain constant, which potentially results in suboptimal energy harvesting during certain times of the day or year [2]. Conversely, tracking systems that adjust according to the sun's position can maximise irradiance capture by maintaining optimal panel orientation throughout the day, which significantly increases energy yield.

Ambient temperature is another critical factor that affects the electrical output of PV modules. As temperature increases, the efficiency of solar cells generally decreases. This is due to the PV effect being negatively correlated with temperature, which suggests that higher temperatures lead to reduced energy conversion efficiency for PV cells. According to the findings of Jordan and Kurtz [3], conventional silicon-based PV systems experience performance degradation with rising temperatures, with efficiency losses ranging from 0.4% to 0.5% for each degree Celsius increase [3]. This temperature sensitivity is particularly pertinent in regions that frequently experience high temperatures, such as South Africa, which might affect the long-term sustainability of PV installations if not accurately modelled and managed.

Wind speed is another significant environmental factor that influences the thermal behaviour of PV systems. Higher wind speeds tend to cool PV modules, which prevents overheating. This efficient cooling can mitigate the adverse effects of elevated temperatures, which may enhance system efficiency. Wind speed plays a prominent role in dual-axis tracking systems, which continuously adjust their orientation to follow the sun's movement. These systems are typically more vulnerable to weather conditions, particularly wind, thus necessitating consideration of wind speed when evaluating system performance. High wind speeds improve heat removal from PV modules, which results in greater overall system efficiency compared to fixed systems, which may not achieve the same level of cooling.

An increasingly significant environmental parameter is soiling, which is defined as the accumulation of dust, dirt, and other particulates on the surface of PV panels. This layer of debris can substantially reduce the amount of solar radiation that reaches the panel surface and, consequently, energy production. In desert and arid climates, such as South Africa's Free State province, the impact of soiling can be a persistent issue, especially for ground-mounted installations that are particularly susceptible to dust accumulation. Under certain conditions, soiling has the potential to cause efficiency losses of up to 30% [4]. Modelling the effects of soiling is therefore essential for achieving reliable long-term performance estimates for PV systems, particularly in semi-arid and desert climates.

Shading is yet another environmental condition that can affect the output of PV systems. Partial shading can lead to significant reductions in energy output due to the interconnection of solar cells in modules that results in "hot spots" that can further decrease efficiency and potentially damage the system. For dual-axis tracking systems, shading impacts can be mitigated more effectively by adjusting panel angles to avoid potential shading from structures or other panels in the array. However, in urban or suburban contexts, where buildings, trees, or other obstacles may cast shadows, shading can significantly influence energy production, particularly during early morning or late afternoon hours.

Geographical location is therefore a critical determinant of a location's solar potential. South Africa benefits from high solar radiation levels due to its advantageous geographical position and climatic conditions, particularly in the northern and central regions. However, despite the high potential for solar energy production, variability in climatic conditions, as well as seasonal fluctuations in solar irradiance and temperature, must be considered in any performance modelling approach. The Free State province, home to the PV plant at the Central University of Technology, Free State (CUT), offers excellent conditions for solar energy production; however, its semi-arid climate, characterised by dust storms and extreme temperature fluctuations, poses challenges for consistent energy generation throughout the year.

The geographical characteristics of a location directly influence the levels of solar irradiance and climatic conditions to which a PV system is exposed. It is imperative that these factors are adequately considered when developing performance models, as a system's efficiency can vary significantly based on installation location. Geographical differences are a primary reason why models that do not account for region-specific climatic conditions may be inadequate for predicting PV performance in a context such as South Africa.

While a substantial body of research has focused on the performance of fixed PV systems, relatively few studies have investigated the unique challenges and performance characteristics of dual-axis tracking systems, particularly in regions with variable climates such as South Africa. Dual-axis tracking systems offer numerous advantages over fixed systems by continuously adjusting the panel's angle to track the sun's movement, which results in energy yields that are typically 35% to 45% higher than those of fixed systems [5]. However, these systems are more complex and susceptible to a broader range of environmental factors, including those previously discussed, which complicates accurate performance assessment.

Current models for PV performance have largely relied on simplifying assumptions that may not accurately reflect the intricate interrelationships among environmental components that impact tracking systems. Many models, for example, often overlook the effects of wind speed, shading, and soiling on dual-axis systems, despite the significant influence these factors can exert on their output. Furthermore, while numerous scholarly investigations have addressed tracking systems in temperate or tropical regions, few have considered the semi-arid conditions that typify South Africa, where dust and temperature variations can severely compromise system performance.

This study aimed to address this gap by developing a comprehensive model that integrated real-world data from the CUT's dual-axis PV plant, including meteorological data, performance data from the Sunny Portal system, and environmental data such as wind speed and temperature. By incorporating these data into advanced machine learning (ML) methodologies, this research sought to create a predictive model that could accurately evaluate the performance of dual-axis PV tracking systems under the specific climatic conditions of South Africa's Free State province.

The primary objective of this study was to address the research gap by developing and validating an extremely accurate and dependable predictive model for the energy output and performance of dual-axis PV tracking systems in South Africa, with specific application to the CUT installation in Bloemfontein. The investigation focused on utilising actual data from the CUT's PV power plant to gain insight into system performance based on local environmental conditions. By leveraging ML techniques, this research aimed to provide energy planners and engineers with a valuable tool for optimising the design, operation, and maintenance of dual-axis PV systems in regions with comparable climatic conditions [6].

1.2 PROBLEM STATEMENT

The increasing demand for clean energy solutions has led to an increase in the utilisation of PV systems. As the global energy sector transitions towards more sustainable and cleaner sources of energy, reliable predictions of the performance of PV systems become essential for feasibility studies and operational improvements. Accurate predictive models for performance are necessary for assessing the economic viability, design, and long-term operational efficiency of PV systems, particularly for large-scale installations.

Despite significant advancements in solar energy technology, a notable gap in both the literature and practical applications exists, as there are currently no proven comprehensive models designed for predicting the performance of dual-axis PV tracking systems, especially in regions such as South Africa. While numerous fixed PV system models are available, most fail to account for the dynamic behaviour of dual-axis systems, which continuously adjust their orientation throughout the day in response to the movement of the sun. This tracking capability allows such systems to capture more solar irradiance and, consequently, generate approximately 35% to 45% more energy compared to fixed systems [5]. However, the inherent complexity and exposure to varying weather conditions in tracking systems pose challenges for modelling and performance prediction.

Such gaps are of more than theoretical significance and instead have profound practical implications for the energy landscape in South Africa. A poor prediction model for a dual-axis tracking mechanism might give rise to incorrect assessments of viability studies, such that either overly optimistic estimates of energy potential are made or no investment should be made due to pessimistic predictions. This directly gives rise to a scaled-up risk with regard to investments made in such energy projects. A poor prediction model might give rise to poorly planned design and maintenance outputs with regard to a dual-axis PV system. This directly leads to levelised costs that might be more prone to inefficiencies with regard to energy. This might give rise to a nation such as South Africa with lingering energy needs being far more hampered with regard to planning energy in a more accurate manner.

In South Africa, a country endowed with abundant solar resources yet characterised by mixed climatic conditions, PV systems represent a critical solution to energy supply challenges, particularly in rural and underdeveloped areas. Although fixed-tilt PV systems have been increasingly studied and implemented, dual-axis tracking systems, which offer greater potential for enhancing solar energy generation, remain under-researched. This discrepancy is particularly concerning given the growing need for large-scale dual-axis tracking systems in South Africa, which is driven by initiatives such as the Renewable Energy Independent Power Producer Procurement Programme (REIPPPP).

Existing performance models for PV systems predominantly focus on fixed systems and are often constrained by simplifications that do not adequately consider the dynamic environmental parameters relevant to dual-axis systems, such as wind speed, temperature fluctuations, and instantaneous variations in irradiance. Many prediction models, including those employed by software like Renewable Energy Project Analysis Software (RETScreen) and Photovoltaic System Simulation Software (PVsyst), assume ideal solar radiation conditions, temperatures, and other parameters while neglecting local climatic variations and transient phenomena such as dust accumulation or shading [7]. These programmes may provide a general overview; however, they lack the requisite detail to accurately represent the behaviour of tracking systems, particularly under variable conditions encountered in regions such as the Free State province of South Africa, where dust, wind, and high temperatures are significant factors.

Moreover, South Africa's distinctive climatic conditions, characterised by sunny days and cold, dusty environments, pose specific challenges for PV installations. Of particular concern is the soiling of panels due to dust accumulation and the thermal changes that occur throughout the day and across seasons, which can significantly reduce energy output if not thoroughly considered. This aspect is frequently underrepresented or overly generalised in models, which leads to overestimations of energy production and a minimisation of maintenance requirements and energy losses over the long term.

The lack of an accurate, real-time prediction model for dual-axis systems that integrates these environmental parameters creates a substantial gap in understanding the cost-effectiveness and long-term performance of such systems. This gap hampers the ability of energy planners, investors, and policymakers to make informed decisions regarding the deployment, scaling, and optimisation of South African PV technology, particularly in rural and off-grid areas that stand to benefit the most from solar power solutions.

Accurate predictions of large-scale dual-axis PV tracking systems are crucial for several reasons. Firstly, such systems represent significant capital and resource investments, and without a reliable prediction model, investors cannot ascertain the economic returns on their investments. Energy output predictions inform investment decisions based on calculations of return on investment (ROI) and the levelised cost of electricity (LCOE), among other factors that are critical for appraising the financial viability of energy projects [8]. A tool that accurately predicts energy output across diverse environmental conditions will provide stakeholders with the necessary information to evaluate economic viability and make strategic investment decisions regarding solar technology.

Secondly, system optimisation and operational efficiency are closely linked to accurate modelling. Dual-axis trackers, while more effective in terms of energy output, require more frequent maintenance and monitoring due to their complexity [5]. Accurate performance forecasting can aid in system optimisation by minimising downtime and identifying areas of potential performance loss that is attributable to environmental factors such as dust accumulation or extreme temperatures [9]. Without a detailed model that accounts for the dynamic nature of the system, operators risk inefficient resource utilisation and potential capital losses.

Moreover, this study sought to address the need for a model that integrates real-world data obtained from operational systems in order to facilitate a more evidence-based and grounded approach to performance forecasting. Current research predominantly relies on models developed through simulations or laboratory conditions, which may not accurately reflect the complexities of outdoor operations [3]. Utilising real-world data from the CUT's PV system in South Africa's Free State province, this research aimed to construct a model that could accurately reflect actual operational conditions, thereby enhancing the reliability and practical applicability of predictions. Real-time data from the Sunny Portal monitoring system were incorporated into the model to account for varying environmental parameters, thereby strengthening its predictive capabilities.

This study endeavoured to address a gap in the literature by developing and testing a comprehensive model that has the potential to forecast the performance of dual-axis PV tracking systems, specifically tailored to South African conditions. By leveraging real-world data from the CUT's PV array, including meteorological data, environmental conditions (such as wind speed and temperature), and energy output data, this research contributes novel insights by providing a model that predicts energy output. Furthermore, it incorporated sensitivity analysis to evaluate the influence of various parameters, including the effects of wind speed and ambient temperature, on system efficiency.

The research employed advanced ML techniques, such as support vector machines (SVMs) and ensemble trees, to create a highly reliable performance forecasting tool. These methodologies facilitated the modelling of complex and non-linear relationships between system output and environmental parameters, which are inadequately addressed by conventional modelling approaches [4]. The results provide energy planners and stakeholders with a detailed tool for energy planning, which offers insight into the operational and financial feasibility of dual-axis PV tracking systems in South Africa.

By addressing this knowledge gap and delivering a validated prediction model, this research makes a valuable contribution to understanding dual-axis PV system performance under real-world conditions. Furthermore, it serves as a useful framework for future large-scale PV applications in South Africa by equipping stakeholders with the tools necessary to make informed decisions regarding technology adoption, performance optimisation, and long-term energy planning.

1.3 RESEARCH AIM AND OBJECTIVES

The primary objective of this study was to address the research gap by developing and validating an extremely accurate and dependable predictive model for the energy output and performance of dual-axis PV tracking systems in South Africa, with specific application to the CUT installation in Bloemfontein. The model incorporated all relevant environmental and operational parameters that influence the performance of PV systems, including solar irradiance, temperature, wind speed, relative humidity, and other local climatic factors. Given that South Africa is characterised by semi-arid conditions and variable weather patterns, there is a need for a model that can effectively simulate the dynamic behaviour of dual-axis tracking systems and provide a dependable forecast of their energy output for the future.

To achieve this primary objective, the following specific objectives were identified:

- Conduct a comprehensive literature review: This objective involved an extensive exploration of the existing body of work pertaining to the determinants of PV system performance, encompassing both environmental conditions (such as temperature, wind, and solar radiation) and the diverse modelling methodologies employed for predicting PV performance. This review had to provide a robust theoretical foundation for the development of the performance model and identify gaps in the literature, particularly concerning dual-axis tracking systems.
- Collect and analyse real-world performance data: The research gathered empirical performance data from dual-axis PV systems currently operational in the Free State province, specifically from the PV plant at the CUT. The collected data encompassed environmental conditions (including ambient temperature, wind speed, and solar radiation), as well as power output from the PV systems, which were utilised for model construction and testing.
- Construct a mathematical model: The study involved the development of a mathematical model that could accurately represent the operational characteristics of dual-axis PV tracking arrays. The model incorporated the influence of significant environmental parameters and the variable characteristics of tracking systems, which enabled it to generate energy output predictions under fluctuating conditions.
- Simulate and validate the model: The model was simulated using real-world data from the CUT's dual-axis PV plant. Validation was conducted by comparing the predicted performance with actual energy output, which allowed for a thorough assessment of the model's accuracy and its capacity to reflect genuine operational conditions.
- Compare simulation output with actual performance data: The accuracy of the model was evaluated through a comparison of its predictions with actual performance data. This assessment facilitated the calculation of a quantitative error

percentage and elucidated the model's effectiveness in replicating real-world performance, thereby contributing to knowledge-based applications and optimisation improvements in PV systems.

By achieving these objectives, this work aspired to develop a reliable and technically accurate model that can be utilised for other dual-axis PV installations in South Africa and similar regions, thereby enhancing the prediction, optimisation, and scalability of PV system performance.

1.4 RESEARCH METHODOLOGY

This study employed a multifaceted and systematic approach to the development and validation of a highly precise model for calculating the performance of dual-axis PV tracking systems within the South African context. Performance estimation was derived from a reference tracker, and through interpolation, results for the entire facility were computed. The research process encompassed four principal phases: literature review, data collection, model construction, and model validation. Each phase was designed to facilitate a comprehensive understanding of the determinants that influence the performance of PV systems, while developing a model that is both academically robust and practically effective.

1.4.1 Literature review

The first part of this study comprises a comprehensive literature analysis of the existing body of knowledge pertaining to the performance and modelling of PV systems. This literature review concentrated on the identification and examination of various parameters that influence the energy production and efficiency of PV systems, with particular emphasis on dual-axis tracking systems. Among the significant parameters investigated were:

- Mono/Abridged Humidity Index: The cumulative effect of relative humidity and black body radiation temperatures as factors for evaluating system performance [3]. High humidity has also been found in other research to generate significant electrical losses in PV systems [10].
- Solar irradiance: Understanding the role of the different classes of solar irradiance (direct normal irradiance [DNI], global horizontal irradiance [GHI], and diffuse

horizontal irradiance [DHI]) in energy generation. Numerous studies have described how the various classes of irradiance affect the behaviour of PV systems, particularly those equipped with PV system trackers [1].

- Ambient and PV cell temperature: Examining the effect of fluctuating temperatures on the performance of PV modules, including material degradation over the long term. The performance of PV cells is significantly influenced by temperature, with higher ambient temperatures leading to decreased performance [11].
- Module degradation: A system of physical and environmental conditions that contributes to a loss of power-generating efficiency in PV modules, such as dust, ultraviolet (UV) radiation, and ageing [3]. Degradation is a prominent factor in long-term performance modelling for PV systems [9].
- Shading and soiling effect: Considering the impact of shading caused by environmental barriers and the accumulation of dirt and dust (soiling) that reduce the output of PV installations, particularly in arid regions such as the Free State [12].
- Orientation and tilt angle of PV module: A discussion of how the orientation and tilt of PV modules influence their output, especially for dual-axis trackers that adjust both the azimuth and altitude angles to maximise energy harvesting [13].

The motivation for this literature review lay in the aggregation of the existing body of knowledge pertaining to these parameters, as well as the identification of potential gaps in research, particularly concerning dual-axis PV tracking arrays in South Africa. This comprehensive background served to inform the development of the prediction model at a subsequent stage.

1.4.2 Data collection

The second step entailed the collection of empirical performance data from a dual-axis PV power plant located in Bloemfontein, South Africa, specifically the CUT, Free State. Data were collected over an extended period to adequately capture seasonal and environmental variations. The primary environmental parameters to be monitored included the following:

- Ambient air temperature: Temperature fluctuations were critical as they affect the thermal loss of the PV modules and their conversion efficiency [10].

- PV cell temperature: The actual temperatures of the PV cells were also measured, as this directly impacts the electrical output of the module; higher temperatures typically result in decreased efficiency [9].
- Wind speed and direction: Wind plays a vital role in the cooling of PV modules, which significantly contributes to the maintenance of optimal operating temperatures and the prevention of thermal loss [12].
- Solar irradiance (DHI, DNI, and GHI): By measuring various classifications of solar irradiance, a comprehensive analysis of how the incoming solar energy is intercepted by the PV modules can be conducted. Solar radiation is a key factor in the efficacy of solar energy systems [1].
- Solar angles (altitude and azimuth): The altitude and azimuth angles of the sun, which vary depending on the day and year, are essential for determining how the dual-axis tracker tilts the modules to maximise energy yield [13].

In addition to data pertaining to the environment, energy output data were collected from the Sunny Portal online monitoring platform of the CUT, which provides real-time performance metrics of the system. These data served as the target output for the validation and training of the predictive model. The data were formatted and cleansed to ensure uniformity, accuracy, and completeness, which were essential for the credibility of the model.

1.4.3 Model development and validation

One of the objective of this study was to develop a mathematical model for the simulation of dual-axis PV tracking arrays. This model took into account the aforementioned environmental parameters and was constructed to estimate energy yield based on actual conditions.

The development of the model followed the following stages:

1. Mathematical modelling: An inclusive mathematical model was established to describe the operational behaviour of the PV system. The model accounted for the dynamic nature of environmental conditions, such as variations in solar irradiance, temperature, and wind speed, and their influence on the output power of the PV module [12].

2. Training of the model: Based on the data collected from the CUT and other selected locations, ML models, including decision trees, SVMs, and ensemble methods, were employed to train the model. The models were validated and refined using MATLAB's Regression Learner App, which provides robust cross-validation to ensure that the model generalises effectively to new and unseen data [18]. Recent studies related to ML applications in solar energy systems indicate that SVMs and ensemble methods yield high accuracy in predicting system performance [13].
3. Validation of the model: Upon training, the model was validated against real performance data gathered from the Sunny Portal platform. This validation involved a comparison between the model-forecasted energy output and the actual output generated by the PV system. Performance metrics, including root mean square error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2), were utilised to assess the model's accuracy and efficiency in predicting real-world performance [6].
4. Performance indices: Furthermore, the research incorporated an evaluation of system performance based on parameters such as system efficiency, capacity factor, performance ratio, and energy loss. These indices were employed to determine the capacity of the model not only to simulate energy output but also to assess operational performance over a specified period [7].

With this process, the model was validated and improved, thereby accurately representing the operational behaviour of dual-axis PV trackers. The final validated model served as an effective tool for forecasting the performance of PV systems under various weather conditions and would be highly beneficial for energy planners, engineers, and policy administrators involved in the design and management of both South African and international solar energy systems.

1.5 EXPECTED OUTCOMES AND CONTRIBUTIONS

This study aimed to deliver a comprehensive set of results and contributions that would provide a valuable tool for optimising South African dual-axis PV tracking systems. These results not only contribute to scientific knowledge in the field but also offer feasible solutions that can be employed by energy planners, policymakers, and engineers to enhance the efficiency and sustainability of solar energy systems. The major expected outcomes are as follows:

1. **A mathematically validated model of high accuracy for dual-axis PV tracking:** The primary deliverable from this research is a high-accuracy prediction model for dual-axis PV tracking. This model was constructed from actual data collected from the CUT's PV system in the Free State. By incorporating environmental factors such as solar irradiance, wind speed, temperature, and humidity, this model was able to simulate dual-axis PV systems under variable conditions from a simulation perspective [3]. It is anticipated that this model will enhance prediction accuracy beyond that attainable by traditional models for fixed PV systems by accounting for the dynamic nature of dual-axis tracking.
2. **An assessment framework for dual-axis PV system technical and economic feasibility:** Another significant outcome is the development of a detailed framework that evaluates the technical and economic viability of dual-axis PV systems within the South African context. This framework provides a systematic approach to evaluating performance parameters such as energy output, efficiency, and the impact of environmental conditions on system performance. Additionally, it encompasses economic parameters such as ROI, payback periods, and cost-benefit ratios, which are crucial in determining the long-term sustainability of large-scale dual-axis tracking systems [8]. The framework can empower investors, policymakers, and energy planners to make informed decisions regarding the implementation of dual-axis PV technology in South Africa's energy sector.
3. **Optimal design, operation, and maintenance guidelines for grid-connected dual-axis PV plants:** The study provides comprehensive guidelines for the design, operation, and maintenance of dual-axis PV systems. These guidelines were informed by the findings from the model and offer recommendations for optimal system configuration, maintenance intervals, and performance monitoring practices. The influence of environmental conditions such as temperature, humidity, and wind speed, which significantly affect the efficiency of a PV system, is emphasised [10]–[12]. Furthermore, the guidelines propose design enhancements to mitigate the degradation effects caused by environmental factors such as soiling, shading, and temperature fluctuations.

4. **Suggestions for enhancing the performance of South African PV tracking systems:** The research provides practical guidelines for improving the performance of operational PV tracking systems in South Africa. These guidelines focus on optimising system efficacy and prolonging the operational life of PV modules by addressing issues such as dust accumulation, panel misalignment, and environmental degradation [11]. The findings from the model can be applied to optimise performance in both operational and newly constructed PV farms in South Africa, thereby achieving higher energy production and enhanced system durability.

1.5.1 Data collection and preprocessing

To be considered for inclusion in the study, a case study was conducted based on the Free State PV power plant, utilising data from meteorological stations and the Sunny Portal monitoring system at the CUT. The parameters examined included fundamental environmental factors such as ambient air temperature, solar radiation, wind speed, and the output power of the PV system. Data cleaning and preprocessing were performed using MATLAB to ensure the reliability and consistency of the dataset. This step was crucial, as it rectified any gaps or inconsistencies in the data that could compromise the accuracy of the predictive model [19].

1.5.2 Model development

To develop the predictive model, this study employed multiple regression techniques, including:

- Decision trees: Decision trees were utilised for modelling complex non-linear relationships between environmental parameters and PV output power, thereby facilitating a transparent and comprehensible decision structure [7].
- Regression trees: Regression trees were employed to model the relationship between environmental conditions and PV output. This non-linear modelling technique has demonstrated strong performance in related applications and was assessed for its capacity to forecast PV behaviour under varying conditions [11].
- Neural networks: Artificial neural networks (ANNs) were employed to learn from the data with the aim of uncovering intricate patterns and associations among

environmental constituents that may not be readily identified by traditional regression models [13].

Both validation and training employed techniques such as five-fold cross-validation to develop more generalised models that are capable of performing well not only on the training dataset but also on new data. This technique minimised overfitting and maximised the model's generalisability for real-world applications [6].

1.5.3 Model evaluation metrics

The model's performance was assessed through a few prominent metrics:

- RMSE was employed to ascertain the predictive accuracy of the model by calculating the discrepancies between the actual and predicted output energy [6].
- MAE offered a straightforward indication of the average magnitude of error in predictions, irrespective of their direction [7].
- R^2 was utilised to assess the degree to which the model accounted for the explainable variability of the observed data, with a result close to 1 indicating a more substantial fit [6].

These evaluation criteria were essential for ascertaining the reliability and accuracy of the developed model, thereby qualifying it for real-world forecasting and decision making.

1.5.4 Scientific contribution

This research also contributes to the establishment of more accurate models for forecasting the performance of dual-axis PV tracking systems, specifically within the South African context. By enhancing the understanding of how environmental conditions influence the performance of PV systems, this research facilitates the ability to accurately model energy output across a variety of climatic conditions. Additionally, this study provides a comparative analysis of simulated versus actual data that yields valuable feedback regarding areas for improvement and the limitations of current modelling approaches [3].

1.5.5 Social and economic impact

The findings of this research are anticipated to generate significant social and economic outcomes. By optimising dual-axis PV tracking systems, the research enables higher energy yields, thereby transforming solar power into a more reliable and cost-effective energy source in South Africa. The model developed in this research assists in predicting energy output, which can facilitate the integration of PV systems into the national grid and contribute to grid stability. The economic implications of this research provide essential insights for enhancing the performance of currently operational PV plants and informing future investments in new renewable energy infrastructures. Furthermore, the research supports South Africa's renewable energy policies that are aimed at promoting sustainability and reducing reliance on fossil fuels [8]. In addition to the expected outcomes described above, the study produced several concrete research contributions, which are outlined in the following section.

1.6 RESEARCH CONTRIBUTIONS

This study presents several groundbreaking contributions to the development of PV energy systems, particularly in emerging economies such as South Africa, where energy demand, grid power reliability, and sustainability objectives intersect. Firstly, the study introduced a robust ML predictive model specifically designed for dual-axis tracking systems in semi-arid locations. By incorporating real-time meteorological inputs such as DNI, GHI, temperature, humidity, wind speed, tilt, and azimuth, the model accounted for the non-linear interactions between environmental factors and PV output in complex and dynamic ways. Unlike theoretical models or static regression models, the ensemble algorithm demonstrated a high degree of adaptability and accuracy, achieving an R^2 value of 0.99338, which surpassed that of conventional software such as PVsyst, RETScreen, and SAM. This illustrated the potential of AI in bridging the gap between theoretical and actual performance.

Secondly, this study offered a validated methodology for PV system performance analysis using actual operational data from the CUT. The integration of data preprocessing, k-fold cross-validation, model benchmarking, and performance visualisation provided a replicable framework applicable to peer institutions or municipalities seeking to monitor and enhance solar assets. The rigorous application of software such as MATLAB's Regression Learner App and the Sunny Portal monitoring system illustrated the ability of affordable digital tools to enhance technical capacity without reliance on costly proprietary software. Moreover, the incorporation of degradation analysis, evaluated against manufacturer ratings and monitored over a five-year period, enhanced the accuracy of long-term energy yield forecasting, which is a crucial factor for planning and budgeting in both the public and private sectors.

Thirdly, the study established the economic and environmental feasibility of investment in solar infrastructure in the education sector. Through a specified cost-benefit analysis, the study assumed a five-year minimum acceptable rate of return of 2.22%, achieving cumulative savings of R714 000 and estimating long-term savings of R3.48 million over a span of 25 years. These results present a compelling case for decision makers who are considering similar investment options, particularly in the context of rising grid tariffs and unstable power supply in South Africa. Furthermore, the study approximated that the CUT's system avoids the emission of 1.6 million kg of CO_2 annually, amounting to over 40 million kg of emissions avoided throughout its operational lifetime. These savings align with the sustainability parameters set forth in South Africa's IRP 2019 and bolster national ambitions under the United Nations Sustainable Development Goals (SDGs), specifically SDG 7 (Affordable and Clean Energy) and SDG 13 (Climate Action) [1],[2].

Lastly, the integration of AI with degradation modelling and cost forecasting represents a distinctive interdisciplinary strategy. Unlike most PV studies that focus solely on efficiency or cost analysis, this study amalgamated technical, financial, and environmental considerations into a comprehensive decision support system. The outcome transcended a mere model; it constituted the development of a usable framework for the prediction, planning, and optimisation of the performance of solar systems in practical applications. The emphasis on real-time variability and environmental response also established a new benchmark for future smart grids, particularly in light of South Africa's advancement towards decentralised power and microgrid applications.

1.7 LIMITATIONS

While this study aspired to make significant contributions to the modelling and understanding of dual-axis PV tracking systems, certain limitations must be acknowledged. These limitations arise from the scope of the research, the data collected, and the methodologies employed. Despite these constraints, the findings and insights derived from the study offer valuable knowledge that can be leveraged to develop and optimise PV tracking systems, particularly in the South African context.

1.7.1 Focus on model development, simulation, and validation

One notable limitation of this work was that it predominantly centred on model development, simulation, and validation based on the collected data. Although this was a fundamental aspect of the research, it did not account for a broader set of parameters that may influence the ultimate performance and optimisation of PV systems. Specifically, the study did not consider the real-world implementation of the model beyond simulation in an operational environment. Consequently, while the model demonstrated substantial potential accuracy and future applicability based on simulations and validation utilising historical data from the CUT's PV plant, its real-world implementation may encounter challenges that are not addressed in the model.

For example, operational constraints such as maintenance schedules, technical issues with hardware, and unforeseen environmental conditions, such as natural disasters or extreme weather, may impact the actual output of dual-axis PV systems. Furthermore, although the model provides realistic estimates of energy output based on environmental conditions, it lacks mechanisms to incorporate real-time operational adjustments aimed at maximising energy yield, such as tilting and azimuth adjustments of the PV panels based on system feedback or operator input. The primary application of the model is prediction, and while it is grounded in robust theoretical principles, its practical applications in real-world systems may necessitate further verification and fine-tuning based on ongoing data from operational systems.

1.7.2 Geographical limitation of data collection

Another limitation of this research is that data collection was confined to a PV power plant located in the Free State province of South Africa, specifically the CUT. While this focus provided valuable insights into the operational behaviour of PV systems in a semi-arid climate, it is reasonable to assume that environmental conditions, solar irradiance patterns, temperatures, and other parameters may vary significantly in other regions of South Africa or elsewhere in the world.

For example, areas closer to the coast, such as the Western Cape or KwaZulu-Natal, may experience different climatic conditions that could potentially affect the performance of PV systems. Higher humidity levels, increased cloud cover, or greater rainfall might yield different performance outcomes for PV systems compared to the sunny, arid conditions prevalent in the Free State. Consequently, the results of this research may not be universally applicable to other regions that are characterised by markedly different environmental conditions. However, the models and insights generated in this study could serve as a foundational framework for other regions by incorporating local variables such as solar irradiance and weather conditions. Future research could thus extend the geographical scope by collecting data from additional regions in South Africa or even from other countries, thereby enhancing the generalisability and applicability of the model across diverse climatic contexts.

1.7.3 Exclusion of detailed economic cost-benefit analysis

Another limitation of this research is the absence of a comprehensive economic cost-benefit analysis beyond performance verification. Although this work addressed some basic financial indicators, such as ROI and energy production-based savings, it did not account for the full range of economic considerations that may influence the decision to invest in a dual-axis PV tracking system. The economics of such systems are not solely contingent upon their performance and energy savings but also encompass initial installation and maintenance costs, available funding, long-term operational expenses, and existing incentive policies.

Specifically, the upfront costs for dual-axis PV tracking systems are often higher than for fixed-tilt systems due to the increased complexity of the tracking system. While dual-axis systems typically generate more power, particularly under higher solar irradiance conditions, elevated initial costs and potentially higher maintenance expenditure may negate these benefits if not managed effectively. Additionally, the model examined in this study focused solely on energy output predictions and their associated economics. A more thorough cost-benefit analysis would consider lifecycle costs, operational and maintenance expenses, system degradation over time, potential downtimes for repairs or upgrades, and comparisons with other energy sources or technologies.

To realistically evaluate the economic feasibility of dual-axis PV systems, further research could be conducted through a holistic financial model that encompasses the total cost of ownership, including capital expenditures, periodic operational costs, and any government subsidies or incentives. This would facilitate more informed decision making regarding the financial viability of adopting dual-axis PV systems in South Africa and other regions, taking into account both long-term and short-term economic outcomes.

1.8 PUBLICATIONS FROM THIS STUDY

Conference paper:

• Kambolo, MD., Hohne, PA., Kusakana, K. Machine learning-based modeling and validation of dual-axis photovoltaic tracking systems in south Africa. Submitted.

Journal article:

• Kambolo, MD., Hohne, PA., Kusakana, K. Degradation Analysis and Long-Term Performance Forecasting of a Dual-Axis PV Solar Array in South Africa. Submitted.

1.9 CONCLUSION

This chapter outlined the research proposition for developing a forecasting model for dual-axis PV tracking systems in South Africa. In response to the evident gap in the demand for credible performance forecasting tools for dual-axis systems, this research sought to address this gap in modelling, particularly within the South African context, by utilising real-world data and advanced ML techniques. In this endeavour, a representative tracker was employed as a benchmark for data generation, with results extrapolated to provide plant-wide performance figures.

The investigation focused on integrating local environmental conditions, specifically wind speed, temperature, and solar irradiance, whose contributions significantly impact PV output. Validation of the model through data from the CUT yielded useful insights into the unique climatic conditions present in South Africa, which would facilitate optimal system design and operation.

Furthermore, this research contributes a model for assessing the technical and financial viability of commercial-scale PV systems. By forecasting system energy output and degradation, the model enhances maintenance planning and promotes the economic potential of PV systems. Ultimately, this work supports South African renewable energy policy and enhances the effectiveness of solar power integration into the national power grid.

In summary, this research offers a potential tool for maximising power generation from solar energy in South Africa, thereby facilitating the country's transition towards sustainable energy sources and energy security.

CHAPTER 2:

LITERATURE REVIEW

2.1 INTRODUCTION

South Africa is currently grappling with an energy crisis characterised by frequent power shortages, which underscores the necessity for alternative energy sources, particularly solar power. Government initiatives, such as the REIPPPP, have expedited the implementation of large-scale solar energy projects. Notably, dual-axis PV tracking systems have demonstrated energy yields up to 45% higher than those of fixed systems [51], rendering them especially advantageous in regions with high solar potential, such as the Free State.

This chapter provides a critical review of the literature on PV system performance modelling relevant to dual-axis tracking systems. While fixed systems have received substantial attention in the literature, models specifically addressing dual-axis systems are relatively scarce, particularly in the context of South African conditions. The review encompasses a variety of modelling techniques, including empirical, physical, and ML models, and identifies shortcomings in these models regarding their capacity to account for actual environmental variables such as the variability of irradiance, wind, and degradation [47]–[52].

2.2 OVERVIEW OF LARGE-SCALE PHOTOVOLTAIC (PV) SYSTEMS AND TRACKING TECHNOLOGIES

2.2.1 Existing research on large-scale Photovoltaic plants

Research on large-scale PV systems has investigated system design, environmental factors, and performance optimisation. PV power plants are characterised by key components, including silicon PV modules, inverters, support structures, and, increasingly, energy storage systems. The literature indicates that module quality (monocrystalline versus polycrystalline), inverter efficiency, and solar irradiance specific to the locale significantly impact plant productivity [48]. Additionally, performance enhancements are often linked to the installation of tracking systems, with dual-axis models providing superior sun-tracking capabilities [52].

2.2.2 Performance factors identified in the literature

The literature identifies several performance determinants, including solar irradiance, temperature, wind speed, shading, and module degradation. High levels of irradiance, such as those prevalent in South Africa, correlate with increased energy yield; however, elevated temperatures may diminish panel efficiency. Wind can serve both as a cooling source and a contributor to structural loads. Shading and dust accumulation can reduce output by approximately 5% to 20% [50], while long-term module degradation typically ranges from 0.5% to 1% per annum [49]. Consequently, maintenance emerges as a critical factor in ensuring long-term performance success.

2.2.3 Photovoltaic development in South Africa

The South African REIPPPP has been instrumental in the development of utility-scale solar farms, symbolising the country's commitment to clean energy adoption. One of the early frontrunners is the De Aar Solar Power Project, depicted in Figure 2.1, which was commissioned in 2014 as part of the first bid window of the REIPPPP. With a generation capacity of nearly 50 MW and an annual supply of approximately 85 GWh – sufficient to power more than 19 000 homes – the project occupies an area of approximately 100 hectares and was developed by a consortium led by Globeleq and Mainstream Renewable Power, among others.

Additionally, Figure 2.2 shows the Letsatsi Solar Park located in Bloemfontein in the Free State province, which commenced commercial operations in May 2014. This 75 MW plant generates about 150 GWh per year, enough to supply approximately 50 000 to 60 000 households. It represents one of the largest project finance transactions executed in South Africa to date and exemplifies the economic and developmental aspirations that underpin the REIPPPP initiative.

These studies not only document the engineering achievements of commercial-scale, grid-connected PV plants in South Africa but also serve as reference points for evaluating the efficiency of advanced solutions such as dual-axis tracking systems. Their implementation underscores the strategic importance of solar power deployment in addressing the nation's energy security challenges and advancing socio-economic goals through leadership in clean energy.

Figure 0.1: De Aar Solar Power Project solar panels [136]

Figure 0.2: Letsatsi Solar Park, a large-scale PV plant in the Free State [137]

2.3 DUAL-AXIS TRACKING SYSTEMS: A LITERATURE PERSPECTIVE

2.3.1 Tracking technology and solar yield

Studies differentiate between single-axis and dual-axis trackers, with the latter demonstrating superior performance owing to their capacity to adjust panels along both horizontal (azimuth) and vertical (altitude) axes. The literature indicates energy gains ranging from 35% to 45% over fixed systems in regions that are characterised by high irradiance. Table 2.1 summarises yield improvements observed in South Africa and California in the United States of America (USA), which reinforce the efficiency advantage of dual-axis trackers [51].

Table 0.1: Yield improvements for dual-axis trackers

Location	Fixed system yield (kWh/m²)	Dual-axis tracking yield (kWh/m²)	Percentage gain (%)
Free State, South Africa	1 500	2 000	33.33
California, USA	1 800	2 600	44.44

2.3.2 Operational design and control strategies

Advanced dual-axis tracking systems utilise microcontrollers and photoresistor-based sensors for real-time solar tracking. Structural designs vary, but common elements include stepping motors, gear assemblies, and feedback systems to ensure optimal orientation. Control precision is crucial for maximising performance, particularly under varying environmental conditions. Figure 2.3 illustrates the structural diagram of a solar tracking device [140].

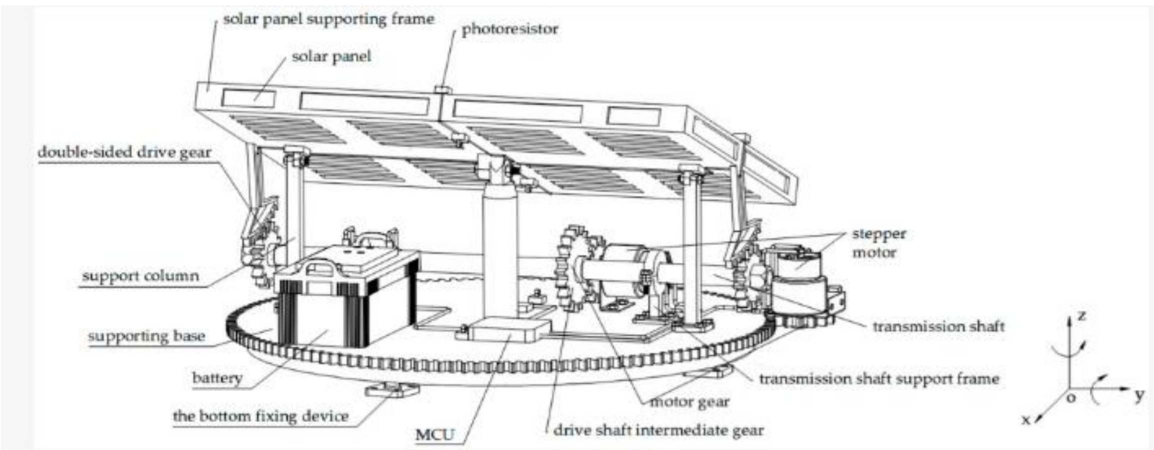

Figure 2.3: Structure diagram of a solar tracking device [140]

2.3.3 Environmental impact and efficiency

The wind speed has a demonstrable cooling effect on PV panels, which enhances performance in warm climates [51]. However, excessive winds pose a mechanical hazard. Dust deposition and selective shading are significant deterrents, with reports highlighting the need for frequent maintenance and cleaning intervals [50]. Furthermore, the literature advocates for the incorporation of inherent self-cleansing and automatic maintenance provisions to prevent losses in efficiency [52].

2.3.4 Cost-benefit analyses in the literature

Despite higher installation and maintenance costs, economic analyses in numerous studies suggest that the increased power output of dual-axis systems justifies the investment in regions with greater solar potential [49],[51]. Nonetheless, the outcomes are site-specific, and additional studies are required to enhance economic estimations and long-term system reliability.

2.3.5 Research and future directions

Despite the considerable potential of dual-axis tracking systems in maximising energy yields, further research is necessary to fully comprehend the interactions of environmental factors such as wind speed, temperature, and soiling in locations with distinct climatic conditions. Future studies may involve enhancing the reliability and longevity of the systems, developing more advanced automated maintenance systems, and creating more robust cost-benefit assessment models to render the systems more financially competitive in various locations.

Dual-axis tracking technology represents a significant advancement in solar energy capture, offering up to 45% higher energy yield compared to fixed systems. However, its effectiveness depends on local climatic conditions and system maintenance. In regions such as the Free State, South Africa, where solar irradiance is high, dual-axis trackers can substantially improve energy production; however, careful attention must be paid to environmental factors, system durability, and maintenance. Future advancements in tracking technology, coupled with optimised maintenance strategies, will enhance the viability and efficiency of dual-axis tracking systems, which will contribute to a more sustainable energy future.

2.4 MODELLING TECHNIQUES IN PV SYSTEMS

A range of modelling methods has been developed to predict PV system output tailored to specific objectives and operational contexts. These models are valuable for maximising system design, estimating energy yield, and supporting predictive maintenance. Broadly speaking, PV output modelling methods are categorised into empirical models, physical models, and ML models. Each type has distinct strengths and weaknesses concerning complexity, environmental variability, and data support. Empirical models are statistically derived from data and incorporate environmental inputs and corresponding energy outputs in a data-driven approach. Physical models simulate PV system performance based on the fundamental principles of the physics of conversion processes, taking into account variables such as irradiance, temperature, and module properties. ML models extract non-linear patterns and interactions of variables from historical data and offer flexibility in highly dynamic environments.

The selection of the appropriate modelling approach depends on the type of deployment location, access to meteorological data, computing power, and the intended application, whether for system simulation, forecasting, or performance optimisation. While empirical models may suffice for initial feasibility studies that require rapid approximations, physical models are more suitable for rigorous engineering design. ML models are increasingly favoured in smart grid applications, where near-real-time forecasting and adaptive learning are crucial. Furthermore, hybrid models that integrate two or more methods are gaining traction, with the aim of enhancing accuracy, robustness, and adaptability. These models leverage the physical interpretability of physics-based methods and the predictive capabilities of data-driven algorithms. They represent a promising future direction for PV system modelling and forecasting applications, particularly in highly complex and variable environments such as South Africa.

2.4.1 Empirical models

Empirical models are among the earliest and most frequently employed predictive models for PV system performance. These models utilise historical data to establish mathematical correlations between output variables (e.g., direct current power or energy yield) and input variables (e.g., GHI, ambient temperature, wind speed, and relative humidity). Their simplicity renders them useful for initial feasibility studies, low-infrastructure sites, and rapid diagnostic analyses [56],[57]. Empirical models often employ statistical techniques such as linear regression, multiple linear regression (MLR), exponential functions, and polynomial fitting.

The earliest empirical model for solar-energy prediction is the Hottel–Whillier model, proposed originally in the classical works of Hottel and Whillier [147] for flat-plate thermal collectors, and later modified by Bliss [148]. Despite being developed for thermal systems, the linear version of this expression has been used extensively in PV simulations. A simplified version according to this definition is used by some authors (e.g., [53]):

$$P = a + bI + cT + dW \tag{2.1}$$

Where:

- P is the predicted power output (kW);
- I is solar irradiance (W/m^2);
- T is ambient temperature (°C);
- W is wind speed (m/s); and
- a, b, c, d are model coefficients determined via regression analysis.

This type of model is particularly advantageous in data-scarce contexts or during the feasibility study phase of a project, where more complex models cannot be implemented. These linear models may be deemed acceptable in the Gauteng and Northern Cape provinces, where it has been established that they yield satisfactory levels of error – within ±10% on a month-to-month basis – when calibrated using regional climatic data [59].

Empirical models are inherently constrained from a statistical perspective, as they depend on fixed relationships that do not necessarily account for non-linear interactions, lag effects, or system degradation. Consequently, their accuracy is diminished in situations that are characterised by variable cloud cover, dust deposition, or shifting orientation and shading effects. These limitations are particularly pronounced in the context of high atmospheric variability observed in sub-Saharan climates.

To address these deficiencies, researchers have proposed advanced empirical methods, such as:

- Stepwise regression: This method automatically selects the most statistically significant variables, which mitigates the risk of model overfitting.
- Autoregressive integrated moving average (ARIMA): This approach preserves time series dependencies and demonstrates efficiency in short-term forecasting under stable conditions [58].
- Hybrid support methods: While empirical models provide initial estimations, ML techniques are employed for subsequent correction layers [60].

A prominent study conducted by Mellit and Kalogirou employed MLR to analyse hourly PV power generation in Mediterranean climates, utilising factors such as irradiance, temperature, and humidity, and achieving R^2 values of up to 0.92 [15]. However, when this methodology was applied to South African data, the results were less satisfactory, primarily due to heightened variability in irradiance levels and the presence of atmospheric dust, which were not adequately accounted for in the empirical models.

Table 2.2 provides a summary of common empirical model forms, input variables, and typical performance metrics.

Table 2.2: Common empirical models for PV output forecasting

Model type	Input variables	Region of application	R^2 range	MAPE (%)
Linear regression	Irradiance, temperature	South Africa (Free State)	0.85–0.90	6–10
Polynomial regression	Irradiance, temperature, wind	Tunisia, Morocco	0.88–0.91	5–9
MLR + humidity	Irradiance, temperature, humidity	Jordan; South Africa	0.82–0.88	7–12
ARIMA	Lagged irradiance	Spain, Egypt	0.87–0.92	5–8

MAPE = Mean absolute percentage error

Another significant disadvantage is the lack of physical interpretability; whereas physics-based models provide insight into how individual system components or degradation mechanisms affect performance, empirical models do not offer comparable information. This absence of explainability presents challenges in predictive maintenance or diagnosis, where a comprehensive understanding of failure modes is essential.

Recent developments aim to integrate empirical models with dynamic sensor networks in real time by utilising streaming data on module temperature and irradiance to adjust the coefficients dynamically. Nevertheless, this approach remains inadequate in the face of sudden weather changes (e.g., moving clouds) and leads to errors that static regression parameters are unable to accommodate. The empirical model residuals depicted in Figure 2.4 demonstrate a marked increase under unstable weather conditions, primarily because dynamic situations render the residuals less reliable.

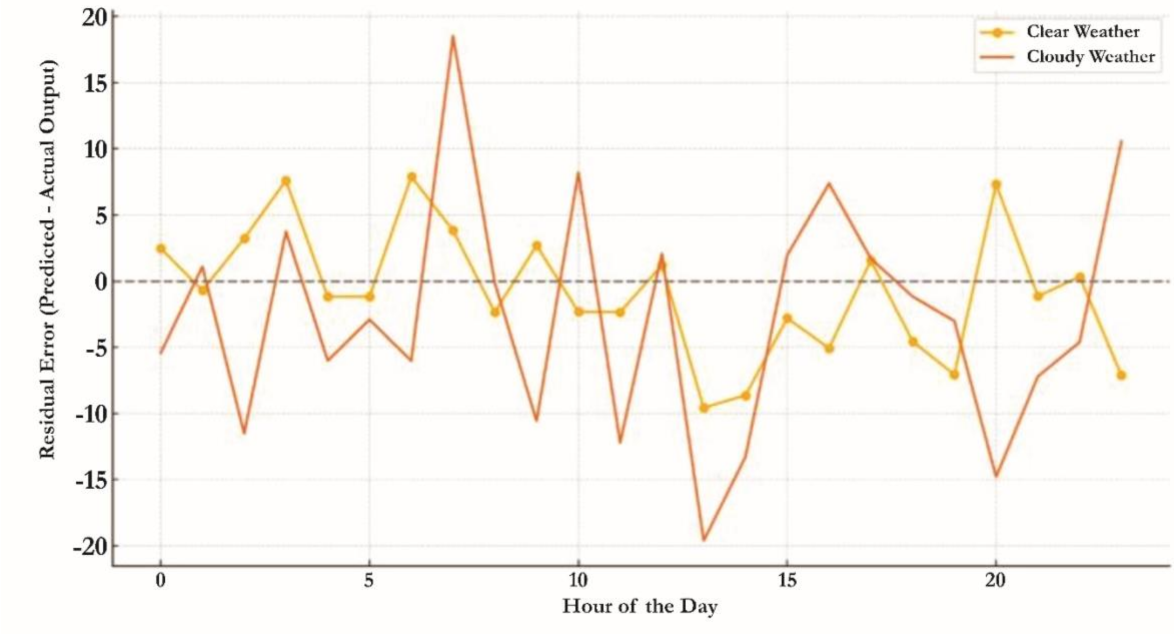

Figure 2.4: Residual errors of empirical PV output models under variable weather conditions

Empirical models exhibit simplicity and ease of implementation; however, their performance is significantly limited by environmental variability and the lack of underlying physical mechanisms. Their application is most suitable for initial examinations or in contexts where the efficiency of the computational code is emphasised, often at the expense of precision. The integration of these models with adaptive or hybrid approaches represents an area of ongoing research, which is particularly relevant in regions characterised by high inter-day variability of solar resources, such as South Africa.

2.4.2 Physical models

Physical models provide a mechanistic and high-fidelity simulation of PV system behaviour. In contrast to empirical models, which depend on historical-statistical correlations, physical models simulate the internal thermal and electrical behaviour of PV cells through the principles of semiconductor physics and circuit theory [61]. These models are frequently articulated through equivalent circuit representations, the most recognised of which are the single-diode model (SDM) and the more sophisticated two-diode model (TDM). These models aim to accurately describe the current-voltage and power-voltage characteristics of PV modules under a variety of environmental conditions, including solar radiation, cell temperature, and wind speed [55].

2.4.2.1 Single-diode model (SDM)

The SDM has become the standard characterisation for crystalline silicon PV modules. It comprises a photocurrent source in parallel with a diode, along with a series resistor to represent wiring and contact losses, and a shunt resistor to account for leakage currents. The equation governing the SDM is:

$$I = I_{ph} - I_0\left(e^{\frac{q(V+IR_s)}{nkT}} - 1\right) - \frac{V + IR_s}{R_{sh}} \tag{2.2}$$

Where:

- I : Output current (A)
- V : Terminal voltage (V)
- I_{ph}: Photogenerated current (A)
- I_0: Diode reverse saturation current (A)
- R_s: Series resistance (Ω)
- R_{sh}: Shunt resistance (Ω)
- n: Ideality factor
- q: Electron charge (1.602×10^{-19} C)
- k: Boltzmann constant (1.381×10^{-23} J/K)
- T: Cell temperature (K)

This model strikes a balance between computational simplicity and adequate accuracy across a diverse array of applications. Nevertheless, its performance is compromised under conditions of low lighting or elevated temperatures.

2.4.2.2 Two-diode model (TDM)

To enhance precision, particularly under conditions of partial shading or non-uniform radiation, the TDM incorporates an additional diode to account for recombination losses in the depletion region. While this modification improves the accuracy of the model, it also increases computational costs; consequently, the TDM is better suited for simulation scenarios where a higher degree of precision is necessary [62].

2.4.2.3 Parameterisation and software integration

PVsyst, the System Advisor Model (SAM) developed by the National Renewable Energy Laboratory (NREL), and HelioScope are among the most commonly utilised software tools for the implementation of these physical models. These software applications necessitate comprehensive input data, which include the following details:

- Geolocation and tilt angle;
- Module technology and material parameters;
- Cell temperature coefficients; and
- Meteorological information (hourly or sub-hourly).

These software models simulate system degradation, performance ratios, shading losses, and energy yield. The SAM is particularly preferred in the techno-economic evaluation of systems due to its integration of financial models and energy simulations [63].

2.4.2.4 Key parameters that affect physical models

Table 2.3 summarises the essential parameters of physical models and their corresponding effects on PV performance.

Table 2.3: Essential parameters in physical models and their corresponding effects on PV performance

Item	Parameter	Description	Effect on performance
1	Diode ideality factor	Controls recombination losses in the PV cell	Affects fill factor and efficiency
2	Series resistance	Resists current flow through the cell	Reduces voltage and power output
3	Shunt resistance	Accounts for leakage paths in the cell	Impacts performance at low irradiance
4	Temperature coefficient	Measures sensitivity to heat	Impacts voltage and current generation

2.4.2.5 Applications and case studies

In South Africa, the utilisation of physical models has proven to be instrumental in regions such as the Free State and Northern Cape, where high irradiance and temperature variations necessitate the precise simulation of thermal effects. Research has demonstrated that incorporating temperature coefficients and live module data can enhance the accuracy of the model by 5% to 10% when compared to static empirical models [64].

2.4.2.6 Adaptive and real-time physical modelling

Recent advancements have led to the development of adaptive physical models that automatically adjust model parameters using real-time environmental data sourced from weather stations or Internet of Things (IoT) sensors. Unlike fixed models, adaptive systems can modify the series/shunt resistance and ideality factor in response to variations in irradiance and temperature [65].

An implementation by Liu et al. in semi-arid climates achieved a 13.6% improvement in forecast accuracy by employing real-time updates to the parameters of the SDM [66].

Table 2.4: Comparison of performance metrics between conventional static and adaptive physical models applied in South African conditions

Location	Model type	RMSE (kWh)	Yield accuracy (%)	Error reduction (%)
Free State	Static SDM	1.42	89.2	—
Free State	Adaptive SDM	1.09	93.1	23.2
Northern Cape	Static TDM	1.35	90.3	—
Northern Cape	Adaptive TDM	1.00	94.4	25.9

2.4.2.7 Parameter estimation and model calibration

To accurately calibrate physical models of PV modules, the implementation of parameter extraction algorithms is essential. Commonly employed methods include:

- Newton-Raphson iteration
- Genetic algorithm
- Particle swarm optimisation

These methodologies are frequently utilised to ascertain the values of R_s, R_sh, n, I_0, and I_ph from experimental current-voltage curve data [67].

In this work, the interest is in Newton–Raphson (NR), Genetic Algorithms (GAs) and Particle Swarm Optimisation (PSO) since they represent some of the most used and documented tools in the literature for extracting PV parameters [149]-[151]. General reviews about estimation of parameters in solar cells place NR as a typical example of a deterministic, gradient-based approach, while GA and PSO are examples of evolutionary and swarm-based metaheuristics, respectively. They are often used as benchmark algorithms when comparing different algorithms [149]–[150].

Together, these three methods cover the main families of optimization strategies-numerical local methods, and global metaheuristics-without venturing into the full gamut of newer or more specialized algorithms-such as differential evolution, artificial bee colony, grey wolf optimizer, whale optimization, and others-whose detailed treatment would go beyond the scope of this chapter [151]-[152]. Moreover, all three methods are mature, well supported in standard scientific computing environments, and commonly implemented in PV modeling tools, which further justifies their inclusion here [153]-[154].

This dissertation did not perform new numerical benchmarking of parameter-extraction techniques; instead, it draws on published comparative studies. Several sources indicate that NR remains highly effective for local refinement when good initial parameter guesses are available, while GA and PSO generally offer a favorable balance between global search capability, accuracy, and computational effort when compared with other metaheuristics such as simulated annealing, differential evolution or cuckoo search [150]–

[151]–[155]. For example, Rawat (2019) compared NR, PSO and simulated annealing for single-diode parameter estimation, finding NR to achieve the lowest relative power error and the fastest convergence, while PSO outperformed simulated annealing but with greater computational cost [155]. Broader surveys of PV parameter estimation methods reaffirm that GA and PSO remain standard reference algorithms against which many newer or hybrid methods are assessed [149]–[151]. Therefore, though the discussion in this section is not exhaustive, it is complete with respect to the most established and representative optimisation families used for calibrating single-diode and two-diode PV models.

2.4.2.8 Degradation and lifetime simulation

Physical models are essential in the study of long-term energy production and degradation. Module degradation is typically modelled at a rate of 0.5% to 1% per annum, influenced by factors such as material type, encapsulant ageing, and UV exposure [3]. Physical simulation tools take these factors into account to estimate:

- 25-year cumulative production;
- admission rate; and
- LCOE.

Physical models provide detailed, component-level simulations of PV systems based on real-world physics. They are indispensable for accurate design, operational forecasting, and economic planning. With modern advances in sensor feedback and adaptive modelling, their importance continues to grow, particularly in regions such as South Africa, where climatic variability is significant.

2.4.3 Machine learning (ML) models

ML algorithms have revolutionised PV forecasting by enabling the formulation of non-linear interactions between environmental parameters and system output in complex patterns. While empirical models assume relationships specified *a priori* and physical models rely on fundamental semiconductor models, ML models are entirely data-driven and possess the capacity to automatically identify intricate patterns in large datasets. These characteristics render ML particularly well-suited for short-term, intraday, and minute-by-minute forecasting applications in PV systems [54]–[73].

Among the widely utilised ML models for the prediction of solar power are ANNs, SVMs, random forests (RFs), and gradient boosting machines (GBMs). These models have been deployed in various geographical contexts, including South Africa, in an effort to predict PV power under diverse climatic and irradiance scenarios [69].

2.4.3.1 Artificial neural networks (ANNs)

ANNs are inspired by the structure of the brain and are organised in layers: an input layer, zero or more hidden layers, and an output layer. By employing non-linear activation functions and iterative learning through backpropagation, this architecture is particularly effective for predicting PV power output from highly interrelated variables such as solar irradiance, temperature, humidity, and wind speed [70].

Figure 2.5 illustrates a multilayer ANN architecture designed for PV forecasting. Each input node corresponds to a specific environmental parameter, and weights are adjusted during training to minimise prediction error.

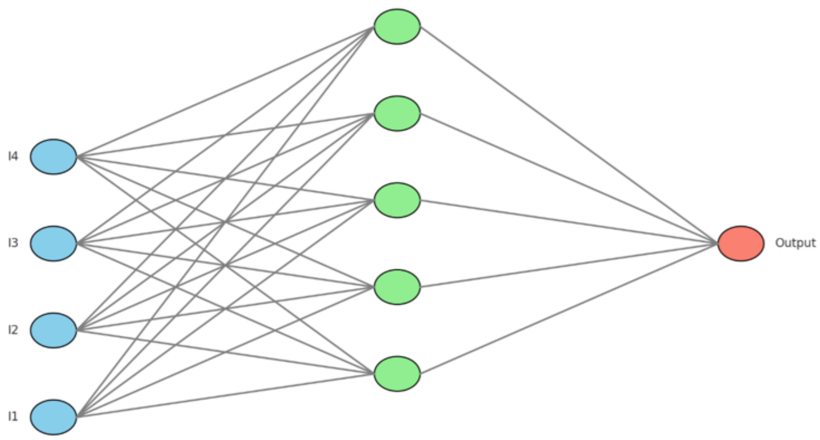

Figure 2.5: Multilayer ANN architecture for PV output prediction [15],[141],[142]

Mellit et al. demonstrated that hourly PV power output could be predicted with an MAPE of less than 5% using ANNs, even in non-static climatic conditions [71]. In the Free State province of South Africa, ANN models that were trained using actual time irradiance and temperature data exhibited superior predictive accuracy compared to empirical models, particularly on partly cloudy days [68].

2.4.3.2 Support vector machines (SVMs)

SVMs are supervised learning models that classify data and perform regression using hyperplanes. Their application in PV forecasting centres on support vector regression (SVR), in which kernel functions such as radial basis function and polynomial kernels enable SVMs to capture non-linear relationships [72].

While ANNs are susceptible to overfitting when trained on small datasets, SVMs demonstrate robustness in high-dimensional and sparse data contexts. For example, Sun et al. employed SVR for the estimation of daily PV production and found the model to be more stable than ANN in the presence of abrupt weather changes [73]. SVMs are particularly advantageous when fewer training examples are available, which makes them relevant for newly commissioned power plants with limited operational histories.

2.4.3.3 Ensemble learning models: Random forests (RFs) and gradient boosting machines (GBMs)

Ensemble methods combine the predictions of multiple base learners to enhance model accuracy and robustness. RFs function by constructing a multitude of decision trees and averaging their outputs, while GBMs train trees sequentially, optimising for error reduction at each stage [16].

RFs offer interpretability benefits, as they provide feature importance rankings, thereby allowing developers to identify the environmental variables that most significantly influence power output. GBMs have been utilised to enhance hourly prediction performance in scenarios characterised by high noise levels [74], particularly with implementations such as XGBoost and LightGBM.

2.4.3.4 Comparative model performance

The performance of these models is contingent upon the data utilised, as well as the variability of weather conditions and the forecasting horizon. Table 2.5 presents a comparative assessment of various ML models regarding their statistical performance in PV output forecasting.

Table 2.5: Comparative performance of ML models in PV forecasting

Model	R^2 score	RMSE (kWh)	MAPE (%)
ANN	0.985	12.4	4.82
SVM	0.979	13.1	5.20
RF	0.981	12.8	4.95
HIMVO-SVM (hybrid)	0.993	9.3	3.67

HIVMO-SVM = Hybrid improved multi-variable optimisation support vector machine

HIMVO-SVM, which integrates heuristic optimisation with SVMs, enhances performance by tuning model parameters through genetic algorithms. Research utilising PV maintenance prediction with HIMVO-SVM has demonstrated R^2 values exceeding 0.99, which indicates its suitability for forecasting and diagnostics [75].

2.4.3.5 Deep learning architectures: Long short-term memory (LSTM) and convolutional neural networks (CNNs)

LSTM networks, a subclass of recurrent neural networks, are designed to accommodate sequential inputs, which renders them appropriate for time series forecasting. LSTMs retain long-term dependencies in memory cells, which enables the precise utilisation of sequential inputs, such as cloud cover and radiation data, to accurately estimate power output [76].

CNNs, while traditionally associated with image processing applications, are also employed in solar forecasting based on satellite images or inputs from sky cameras. For instance, Wu et al. trained a CNN using time-stamped satellite images to predict PV output up to 30 minutes in advance, which achieved enhanced spatial accuracy [77]. Figure 2.6 illustrates a simplified architecture of LSTM employed in PV output forecasting.

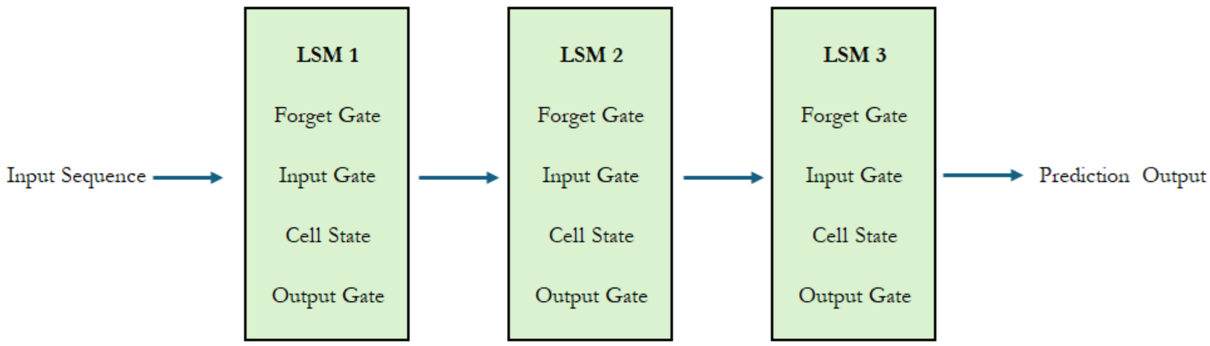

Figure 2.6: LSTM network structure for time series PV forecasting

While these deep learning models frequently outperform classical ML models, they necessitate extensive training datasets, substantial computational power, and meticulous hyperparameter tuning.

2.4.3.6 Challenges and research gaps

Despite the increasing success of ML models in PV forecasting, several challenges persist that hinder their robustness and scalability. One of the primary challenges is overfitting, which is particularly pronounced in ANN models and deep learning models when faced with limited datasets. These models exhibit high precision on training data but fail to generalise adequately to new inputs, which adversely affects their utility [96],[97].

Another significant challenge is the issue of data quality. Real-world PV datasets typically contain missing values, sensor drifts, and noise arising from environmental changes or hardware malfunctions. These deficiencies in data directly decrease the accuracy of predictions and necessitate the application of advanced data cleaning and imputation techniques [98],[99].

Furthermore, generalisability across geographical locations poses a challenge in this domain. ML models trained under specific climatological and topographical conditions often underperform when applied to other locations due to variations in irradiance, humidity, temperature, and microclimate effects [95]. Regional dependence limits the scalability of predictive models unless they are retrained or fine-tuned with regional datasets.

Lastly, the interpretability of complex models, particularly deep learning architectures, remains a critical concern. While these models provide high predictive accuracy, their internal mechanisms are frequently opaque, which complicates efforts to understand or elucidate the rationale behind their outputs. This "black box" nature obstructs transparency, trust, and regulatory approval, especially in mission-critical energy applications [100],[101].

To complement and elucidate the challenges outlined above, Table 2.6 presents a comparative analysis of the strengths and weaknesses of various ML algorithms for PV forecasting, highlighting the trade-offs between them in terms of precision, generalisability, data requirements, and interpretability [95],[98],[99].

Table 2.6: Summary of ML models for PV forecasting

Model	Data requirement	Training time	Interpretability	Ideal use case
ANN	Medium to high	Medium	Low	Hourly forecasting, moderate data
SVM	Low to medium	Low	Medium	Sparse data, regression problems
RF	Medium	Fast	High	Feature analysis, real time
GBM	High	Slow	Medium	High-noise environments
LSTM	Very high	Very Slow	Low	Sequential forecasting
CNN	Very high	High	Medium	Image-based irradiance forecasting

In summary, ML models exhibit unparalleled flexibility and predictive accuracy in PV power forecasting. Their performance is significantly contingent upon the quality of the input data, the selection of appropriate models, and tailored preprocessing techniques. With ongoing advancements in hybrid architectures and data enhancement strategies, ML models are poised to attain greater prominence in the optimisation of PV performance in real-time applications.

2.4.4 Hybrid models

Hybrid models represent a higher tier of modelling that amalgamates multiple approaches – typically empirical, physical, and ML – to leverage the strengths of each while mitigating their respective weaknesses. Such models are particularly advantageous in PV power forecasting, as they facilitate the integration of the interpretability inherent in physical models with the flexibility and precision offered by data-driven methods. In climates akin to that of South Africa, characterised by abundant solar resources and variable weather patterns, hybrid models are increasingly employed to enhance the quality of forecasts and the robustness of systems [79].

A commonly utilised hybridisation technique involves the incorporation of physical parameters into ML models. For example, ANNs may be initialised with values derived from physical models, such as the ideality factor, series resistance, and temperature coefficient. This parameter-seeded learning approach accelerates convergence rates and improves robustness during training, particularly when the volume of training data is limited [80]. By embedding physical behaviour into the training process, hybrid models bridge the gap between theory and observation.

A notable example is the HIMVO-SVM model introduced by Kong et al., which integrates HIMVO with SVR. This proposed model optimised the hyperparameters of SVR – including kernel function type, regularisation coefficient (C), and epsilon-insensitive loss (ε) – using a genetic algorithm, thereby significantly enhancing accuracy levels. In experimental evaluations, the HIMVO-SVM achieved an R^2 value of 0.993 and reduced the RMSE by 28% compared to baseline SVR models [78].

Another method involves the coupling of empirical and ML models to enhance short-term forecasting capabilities. Empirical models provide baseline forecasts through established statistical relationships (e.g., output and irradiance), while the ML models apply a non-linear correction to these forecasts. This two-stage pipeline is particularly effective for day-ahead and intra-hour predictions [81]. Table 2.7 presents a summary of various hybrid model configurations utilised in PV forecasting [54],[98],[102]–[104].

Table 2.7: Common hybrid model configurations in PV forecasting

Hybrid model type	Components combined	Key advantage	Use case
ANN + physical	Physical parameters + neural networks	Faster convergence, improved realism	Short-term output prediction
Empirical + ML	Linear model + RF	Enhanced residual correction	Day-ahead forecasting
HIMVO + SVR	Genetic optimisation + SVR	High accuracy, hyperparameter tuning	Predictive maintenance, anomaly detection
ARIMA + ANN	Time series + deep learning	Captures both trend and volatility	Hourly prediction with historical data

2.4.4.1 Ensemble-based hybrid models

Ensemble learning represents a highly effective paradigm of hybrid modelling. Techniques such as bagging (e.g., RF) and boosting (e.g., gradient boosting, XGBoost) synthesise the predictions of various base learners, thereby reducing variance and the propensity for overfitting. These models are particularly useful in the heterogeneous sub-Saharan African context, where PV performance is influenced by factors such as dust storminess, humidity variation, and frequent changes in irradiance [82]. Figure 2.7 illustrates a simplified ensemble hybrid scheme that integrates different model forecasts (including physical, empirical, and ML models) into a cumulative output, employing a meta-learner such as a linear regressor or neural network.

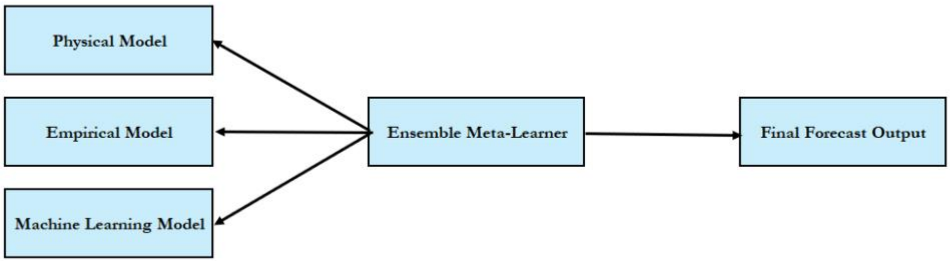

Figure 2.7: Generalised ensemble-based hybrid PV forecasting framework

Boosting models, such as LightGBM and XGBoost, have gained popularity due to their capability to manage missing values, categorical variables, and diverse time intervals. In the study conducted by Zhang et al., the integrated model comprising LightGBM and ANNs demonstrated superior performance compared to individual models for PV output forecasting under varying cloud cover conditions, which achieved an MAPE of 3.3% [83].

2.4.4.2 Physics-informed machine learning (PIML)

The most recent advancement in hybrid modelling is PIML, which integrates physical laws into ML loss functions. For instance, Kirchhoff's current law or equations that govern energy conservation are incorporated into the loss function to regulate the learning process. This methodology enhances generalisation capabilities and improves model extrapolation to novel data scenarios [84].

One notable application of PIML involved the utilisation of irradiance data and physical performance ratios as input constraints to train a CNN, which resulted in an 18% reduction in forecast bias for PV power stations situated in complex terrain [85].

2.4.4.3 Advantages and challenges

Hybrid models serve as powerful tools in PV forecasting by amalgamating various modelling paradigms – primarily physical, empirical, and ML approaches – to harness the strengths of each paradigm. Their comparative advantage is evidenced by improved prediction accuracy. Through the integration of different algorithms, hybrid models mitigate the inherent biases associated with each category of models and yield more robust forecasts across diverse environmental and temporal scenarios [54]–[98]. In addition to offering greater flexibility in exploitation, these models are advantageous in a range of forecasting applications. While physical models are most suitable for long-term strategic energy planning due to their foundation in first-principle physics, ML-driven models excel in short-term, high-frequency PV power forecasting, which enables hybrid systems to facilitate both strategic-level and operation-oriented decision making [95]–[105].

Another observable benefit of hybrid methods is enhanced interpretability. The inclusion of physical or empirical layers renders the forecasting system more transparent, which allows users to trace outputs back to more meaningful input parameters, such as temperature or irradiance. This level of interpretability is crucial in regulated markets where validation and accountability of models are essential [103]. However, despite the advantages they offer, hybrid models are not without their shortcomings. The complexity of such models poses a significant challenge. The integration of multiple forecasting components often necessitates extensive tuning of parameters and prolonged training periods, particularly in contexts that involve deep learning or optimisation algorithms such as particle swarm optimisation or genetic programming [99]–[104].

Furthermore, data synchronisation presents another challenge. The amalgamation of inputs or outputs from various models, especially those operating at differing temporal resolutions, can lead to temporal misalignment issues and an increased likelihood of propagated errors [98]. Lastly, hybrid models frequently demand greater computational resources, which can be a limitation in resource-constrained computing environments, such as edge devices or microcontroller-assisted systems [100]. Consequently, although hybrid models have the potential to revolutionise PV forecasting, their application must be meticulously calibrated to achieve an optimal balance between predictive capability and interpretability, taking into account hardware and deployment-level constraints in real-world scenarios. A comparative analysis of strengths and weaknesses across popular hybrid model paradigms is summarised in Table 2.8 [54],[99],[104],[105].

Table 2.8: Strengths and limitations of hybrid PV forecasting models

Strategy	Strengths	Limitations
ANN + physical	Improves realism and interpretability	Needs initial parameter estimation
HIMVO-SVM	High predictive power	Requires heuristic setup and tuning
Ensemble learning	Robust against noise, scalable	Complex model validation
PIML	Embeds domain knowledge	Still an emerging research area

2.4.4.4 Applications in South Africa

In the South African context, hybrid models have been explored for PV applications in both off-grid and grid-connected scenarios. The notable work by Mthombeni et al. employed a hybrid RF-ANN model to predict PV output in the Free State province. This model adeptly accounted for variability due to dust, temperature, and cloud cover and achieved a yield precision of 92.5% over a test period of three months [86].

In semi-arid regions such as the Northern Cape, where conventional models struggle to predict irradiance due to frequent aerosol occurrences, hybrid models that incorporate satellite imagery (CNN) and sensors demonstrate significant potential for near-real-time output correction [87].

Hybrid models represent a leading edge in PV forecasting innovation by integrating the empirical rigour, physical intuition, and adaptive learning inherent in distinct modelling approaches. As the availability of data and computational resources increases, these models are well-positioned to provide real-time, scalable, and resilient solar performance prediction systems that are tailored to local conditions.

2.4.5 Forecasting tools

Predictive software is a fundamental element in the planning, simulation, and operational optimisation of PV systems. PVsyst, RETScreen, and HOMER Pro are among the most commonly utilised tools, each offering varying levels of capability for both academic and commercial applications. These software products feature graphical user interfaces, standardised irradiance databases, and simulation engines that facilitate the assessment of PV system performance under assumed weather and system conditions [88]. Their user-friendliness promotes rapid feasibility studies and enables stakeholders to conduct economic analyses, system design, and yield estimations.

PVsyst, developed in Switzerland, is particularly popular for the design of grid-tied PV systems. It includes functionalities for shading analysis, performance ratio assessment, and loss factor simulation. The software accepts databases such as Meteonorm or the National Aeronautics and Space Administration's Surface Meteorology and Solar Energy (NASA-SSE) programme for irradiance and temperature profiles. PVsyst simulations rely on deterministic inputs – utilising average or constant meteorological data – wherein inaccuracies may arise under actual conditions where weather patterns deviate significantly from historical records [89]. Figure 2.8 illustrates a typical PVsyst screen used for simulating hourly energy yield in the city of Bloemfontein, South Africa [106].

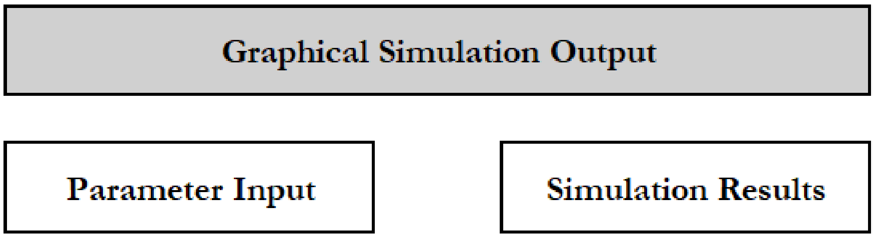

Figure 2.8: PVsyst output interface for hourly yield simulation [106]

RETScreen, developed by Natural Resources Canada, facilitates general evaluations of clean energy projects. It possesses capabilities for financial analysis and emissions reduction tracking in addition to physical PV simulation. Although it is less detailed in its physical PV simulation compared to PVsyst, RETScreen benefits from access to global meteorological databases and the ability to model hybrid systems (e.g., PV + wind + diesel). This makes it particularly convenient for high-level evaluations and comparative studies of feasibility [91].

HOMER Pro, created by the NREL in the USA, focuses on hybrid and off-grid systems. It enables users to optimise configurations based on load profiles, fuel costs, and component efficiencies. For rural electrification or microgrid design in sub-Saharan Africa, HOMER Pro's stochastic simulation engine is especially beneficial. However, similar to PVsyst and RETScreen, it assumes steady-state environmental inputs for each time step and lacks real-time adaptability [91].

To address this static behaviour, recent innovations have introduced ML-based post-processing, whereby outputs from these tools are utilised as initial conditions for training ML models. For instance, irradiance and system loss factors from PVsyst can serve as inputs to ANNs or SVMs that are continuously retrained using weather forecasts and sensor feedback [92]. This hybridisation establishes adaptive forecasting pipelines that correct model drift caused by changes in dust accumulation, temperature variation, or unplanned outages. Table 2.9 presents a summary of the comparative features of the three tools [91],[106],[107].

Table 2.9: Comparative features of major PV forecasting tools

Tool	Strengths	Limitations	Use case
PVsyst	Detailed design, shading analysis, performance ratio calculation	Static inputs, no real-time adaptability	Grid-tied commercial PV plants
RETScreen	Financial analysis, emissions tracking	Limited physical PV modelling	High-level policy or feasibility studies
HOMER Pro	Hybrid/off-grid systems, stochastic simulation	No physical degradation modelling	Microgrids, rural electrification

2.4.5.1 Hybrid forecasting integration

Researchers have developed workflows that integrate PVsyst with data-driven methodologies. For instance, a study conducted by Fadaeenejad et al. involved exporting monthly yield forecasts from PVsyst and subsequently retraining an RF model using actual yield data from a Northern Cape location. The post-processed model enhanced the MAPE from 9.4% to 4.2% [93]. Similarly, NASA-SSE or Copernicus Atmosphere Monitoring Service (CAMS) cloud cover information can be superimposed onto RETScreen results to derive probabilistic yield forecasts with an associated uncertainty estimation. Figure 2.9 illustrates a hybrid forecasting workflow wherein a physical model (PVsyst) trains a neural network that refines its predictions in real time by utilising irradiance and sensor inputs [15],[103],[106].

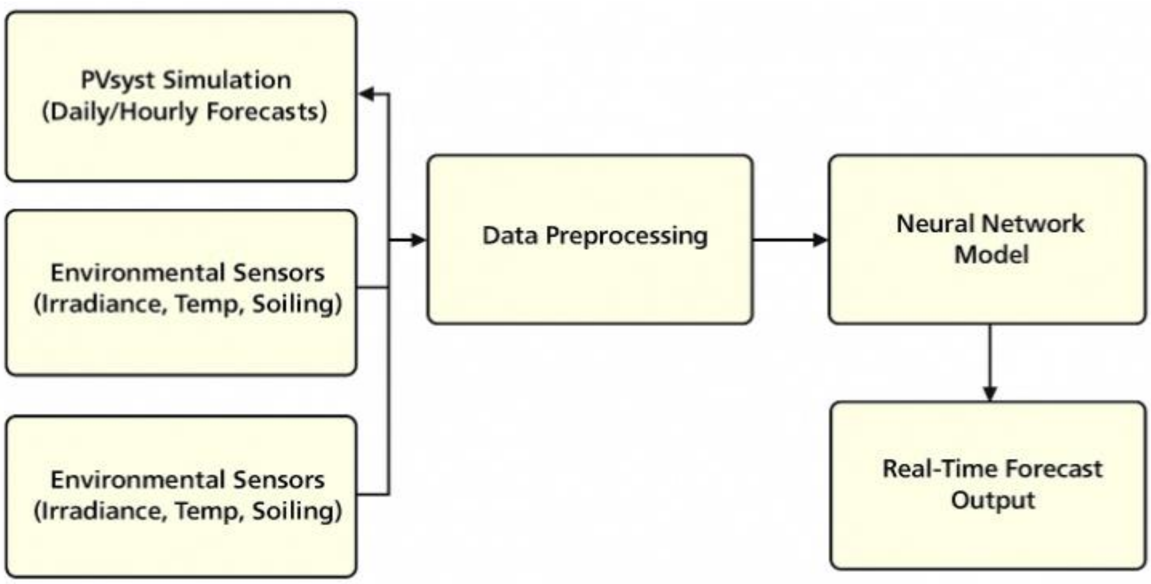

Figure 2.9: Hybrid forecasting workflow combining PVsyst and neural networks

Another encouraging development is the integration of IoT platforms with forecasting software. By connecting devices such as Sunny Portal with ML algorithms, models can automatically import inverter-level measurements (e.g., voltage, current, temperature) and update forecasts without the necessity for manual intervention. These platforms serve as the foundation for true-time digital twins of solar plants – computer models that replicate system behaviour and operation in near real time [94].

2.4.5.2 Advantages and challenges

PV simulation software tools such as PVsyst, HOMER Pro, and RETScreen are widely utilised due to the ease with which simulation inputs can be quickly established in the fields of engineering, finance, and policy applications. These software tools are particularly beneficial in the early-stage feasibility assessments and design of energy systems, owing to their ability to conduct performance ratio analysis, shade simulations, and economic viability assessments [108],[109]. In addition to their primary functions, other software can be incorporated into hybrid ML pipelines, where the deterministic results of the software provide a valuable baseline or feature inputs for more complex data-driven models [110]. This is particularly advantageous in data-scarce locations, such as rural or newly electrified areas, where complete sensor-level data may not be available. As a result, software such as PVsyst serves as a foundation for simulating the baseline levelised energies that can train or calibrate ML models [111].

However, these tools also encounter limitations that constrain their application in advanced forecasting scenarios. Notably, most are not inherently designed for real-time or probabilistic forecasting, which is a capability that is increasingly demanded in modern grid-integrated PV systems, where uncertainty quantification is essential [112]. Furthermore, they rely heavily on accurate, site-specific input data, such as meteorological profiles and system configuration details, which may not always be accessible or reliable [39]. Another limitation is the static nature of their simulation engines; they are not constructed for continuous learning or self-updating based on new data inputs. This renders them less suitable for adaptive forecasting environments, where real-time model retraining is necessary to account for dynamic changes such as panel soiling, inverter degradation, or rapid weather fluctuations [113]. Despite these drawbacks, their strengths render them valuable components in hybrid frameworks and decision support systems when appropriately contextualised.

Table 2.10 illustrates the integration methods and forecasting improvements [108],[110]–[114].

Table 2.10: Integration methods and forecasting improvement

Integration method	Improvement achieved
PVsyst + ANN (retrained)	RMSE reduced by 45% [93]
RETScreen + probabilistic ML	Better cloud anomaly detection
HOMER Pro + SVM hybrid model	20% gain in off-grid reliability

Forecasting tools such as PVsyst, RETScreen, and HOMER Pro are invaluable for PV system design and pre-deployment analysis. However, their static structure limits their effectiveness under rapidly changing environmental conditions. The hybridisation of these tools with ML techniques, including ANNs, RFs, and SVMs, has emerged as a promising approach for developing real-time, site-specific, and adaptive forecasting models. The future of PV forecasting lies in the integration of domain-specific simulation engines with data-driven adaptive algorithms that can respond in real time to environmental dynamics.

2.5 VALIDATION METHODS

Validation is a necessary procedure in the evaluation of predictive models' performance and their capability to generalise in PV output forecasting. Sound validation avoids overfitting and ensures that the model can make reliable predictions of new, unseen data. It is a fundamental component in model construction, particularly in the application of data-driven methods such as ML in solar energy forecasting. Predictive models that are calibrated using specific datasets but do not generalise across varied occasions of time, locations, or weather patterns are considered impractical for applications [17].

Several validation strategies are commonly employed:

- Train-test splits: This method involves dividing the dataset into two segments, typically allocating 70% to 80% for training and 20% to 30% for testing. While this approach is computationally simple and efficient, it may introduce bias if the dataset is not sufficiently large or well-distributed [88].

- Holdout validation: The concept of train-test splits is extended with holdout validation by reserving a separate validation set at the time of training. This technique is particularly useful for tuning hyperparameters in models such as neural networks and GBMs. However, the randomness inherent in a holdout split can lead to inconsistent evaluation results [89].
- Cross-validation: Cross-validation addresses the limitations of the aforementioned methods by enabling the model to be trained and tested on multiple subsets of the data. This approach provides a more robust indicator of performance stability and is widely regarded as best practice in both academic and industrial settings.

The most common metrics evaluated in validation are:

- RMSE: A widely utilised metric that penalises larger errors more severely than smaller ones.
- MAPE: Represents the error as a percentage of the actual value, which facilitates straightforward interpretation.
- MAE: Exhibits reduced sensitivity to outliers in comparison to RMSE.
- R^2: Assesses the proportion of variance accounted for by the model.

Table 2.11: Common validation metrics and their interpretations

Metric	Formula	Interpretation
RMSE	$\sqrt{(\frac{1}{n}\Sigma(y_i\text{-}\hat{y}_i)^\wedge 2\ )}$	Penalises large errors Units match output
MAE	(\frac{1}{n} \sum Y_i - \hat{Y}_i	Average magnitude of errors
MAPE	(\frac {100} {n}\sum\left \frac {Y_i - \hat{Y}_i} {Y_i} \right	Percentage-based accuracy indicator
R^2	$1\text{-}\frac{\Sigma(y_i-\hat{y}_i)^2}{\Sigma(y_i-\bar{Y})^\wedge 2}$	Proportion of variance explained

2.5.1 K-fold cross-validation

K-fold cross-validation is widely regarded by experts as one of the most statistically robust and popular methods for model validation. The dataset is partitioned into equally sized segments, referred to as “folds”. The model is trained on a subset of these folds and evaluated on the remaining fold. This process is repeated multiple times, with each fold serving as the test set only once. The results from each iteration are averaged to provide a final performance metric [91].

Mathematically, the RMSE of each fold is expressed as:

$$RMSE_k = \sqrt{\{\frac{1}{n_k}\sum_{i=1}^{n_k}(y_i - \hat{y}_i)^2\}} \tag{2.3}$$

Where:

- n_k is the number of samples in fold k;
- y_i is the observed output; and
- $\hat{y}_i$ is the predicted output.

The final RMSE is calculated as the average across all folds.

This iterative testing protocol mitigates the variability of the performance metric and provides a more accurate estimation of model behaviour in realistic environmental scenarios. Additionally, it diminishes the risk of overfitting, which is a prevalent issue in ML models, such as neural networks [91].

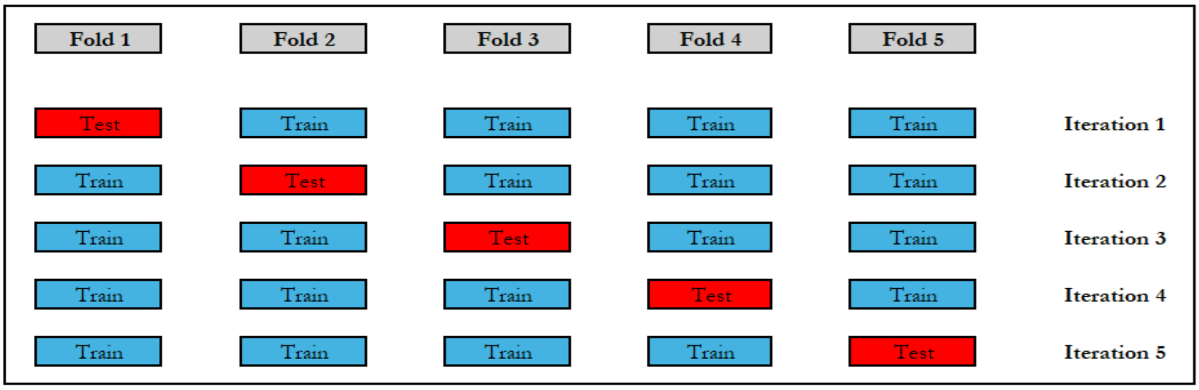

Figure 2.10: K-fold cross-validation workflow for PV model evaluation [143]–[144]

K-fold cross-validation can be customised. K-fold cross-validation is arguably the most widely utilised validation procedure in ML. It provides a robust method for estimating model generalisability across multiple subsets of data. In conventional k-fold cross-validation, the dataset is divided into k balanced subsets (folds), with the model trained on k–1 folds and tested on the remaining fold, iteratively rotating through all partitions. However, conventional k-fold schemes are not universally applicable to all types of data, particularly in classification tasks or time-dependent scenarios.

To address the issue of class imbalance in classification tasks, stratified k-fold cross-validation has been developed. This method ensures that each fold maintains a distribution of class labels that is approximately representative of the entire dataset, thereby mitigating biased validation results, particularly in imbalanced datasets [138],[139]. This approach is particularly beneficial in contexts where the detection of the minority class is of critical importance, such as in fault classification in PV monitoring systems.

For time-dependent data, such as PV power production forecasts, time series k-fold cross-validation presents a more appropriate framework that preserves the sequential order of the data. Unlike random shuffling methods that may disrupt the temporal relationships, this approach refrains from utilising future data to predict past measurements, which maintains the causality necessary for valid forecasting [116]. This methodology is particularly crucial in the context of renewable energy sources, where output performance is significantly influenced by underlying patterns and seasonal variations.

Time-varying PV applications frequently employ rolling-origin or walk-forward validation. These methods uphold the temporal order of the data and simulate real-world forecasting applications by predicting future values based on historical data [92].

Another sophisticated approach is nested cross-validation, particularly for hyperparameter tuning. This technique comprises two loops: the model optimisation inner loop and the model testing outer loop. Although it is computationally intensive, nested cross-validation minimises information leakage and provides a reliable estimate of model performance [93]. Table 2.12 compares various validation techniques in PV forecasting [115]–[117].

Table 2.12: Comparison of validation techniques in PV forecasting

Validation method	Suitable for	Pros	Cons
Train-test split	Large datasets	Fast, simple	High variance, risk of bias
Holdout validation	Medium datasets	Useful for tuning	Depends on split quality
K-fold cross-validation	All-purpose	Low bias, high reliability	Computationally intensive
Time series cross-validation	Sequential PV forecasts	Preserves temporal integrity	Not suitable for random data
Nested cross-validation	Model selection + tuning	Best generalisation, no leakage	Very computationally expensive

In practice, a suite of methods is typically employed to cross-validate PV forecasting models. For instance, an RF model may initially undergo k-fold cross-validation to estimate accuracy, followed by fine-tuning through nested cross-validation to ascertain the optimal hyperparameters [94].

Moreover, the models were also evaluated using live operational datasets derived from monitoring systems such as Supervisory Control and Data Acquisition (SCADA) systems or Sunny Portal systems. These systems provide ground-truth measurements, including voltage, current, irradiance, and module temperature, thereby presenting strong validation scenarios. Such validations account for phenomena such as sensor drift, unanticipated shading, and degradation, which often do not manifest in simulated datasets [95].

In essence, rigorous validation across a range of methods ensures both model reliability and prediction robustness. As the complexity and data density of PV systems increase, validation methods must similarly evolve to accommodate the higher-resolution, multi-source, and multi-modal data landscape.

2.5.2 Holdout method

The holdout approach represents a validation procedure in which the dataset is divided into a test set and a training set. Typically, the dataset is split at a ratio of 70:30 or 80:20, with the model being trained using the training set and subsequently tested with the test set. While the holdout approach is foundational, it may yield less accurate results than cross-validation when the dataset is small.

Performance on the holdout test set is assessed through indicators such as RMSE, MAE, and R^2. This method is more expedient than k-fold cross-validation but is susceptible to significant variation in performance based on the data division.

2.5.3 Comparison with real-time data

In addition to the aforementioned methods, model validation using actual field performance data constituted a critical component of the validation process. Sunny Portal data, a well-regarded online PV plant monitoring system, is frequently employed for this purpose. This approach enables a direct comparison between predicted outcomes and actual output across various environmental conditions [13].

The deviation between actual and predicted production is typically quantified using the following indicators:

- **RMSE**

$$RMSE = \sqrt{\{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2\}} \quad (2.4)$$

Where:

- N is the number of data points;
- y_i is the actual measured output; and
- $\hat{y}_i$ is the predicted output.

- **MAE**

$$MAE = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \quad (2.5)$$

Where the absolute differences between the predicted and actual outputs are averaged.

- **R²**

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \quad (2.6)$$

Where:

- $\bar{y}$ is the actual values' mean; and
- R^2 is the percentage of the observed data variance that is predictable from the model.

These measurements underpin the predictive capability of the model, with lower RMSE and MAE values indicating superior performance, and an R^2 value approaching 1 signifying a greater correspondence between the anticipated and actual outputs [13].

2.5.4 High-quality datasets

The majority of models, including those for PV output, require high-quality datasets to ensure reliability and efficiency in their performance. This is particularly critical when environmental parameters such as solar irradiance, temperature, and wind speed are considered as input variables. The accuracy of the model's forecasts is directly dependent on the quality of the datasets employed in its training [18].

To prevent the model from being adversely affected by data deficiencies, it is essential to clean the data and eliminate any discrepancies, errors, or anomalies. This process may involve the imputation of missing values, correction of inaccuracies in sensor measurements, or remedial actions to address data drift that accumulates over time. High-quality datasets further enable the model to discern the relationships between the input variables (such as solar radiation and temperature) and the output (power generation), which enhances the precision of the validation process [13].

2.6 CONCLUSION

Despite significant advancements in the simulation of PV systems, the understanding of the performance of dual-axis tracking systems remains markedly limited, particularly in regions such as South Africa. While previous studies have predominantly addressed the efficiency of fixed and single-axis tracking PV systems, the adoption of dual-axis tracking technology in locations with specific climatic conditions, such as South Africa, has not been extensively researched. Existing research has largely focused on ideal or controlled climatic environments, which may not adequately represent in situ conditions in semi-arid or variable climates. This knowledge gap is particularly pertinent in light of the ongoing energy crisis in South Africa, where optimising renewable energy sources, such as solar power, is critical for maintaining a stable grid and meeting the nation's energy demands [19].

One notable area that has received limited attention in prior research is the impact of environmental factors, such as wind direction and speed, on PV system output. Although the effects of temperature on PV module efficiency have been thoroughly examined, the role of wind speed in mediating PV cell temperature and enhancing the cooling effect has been insufficiently integrated into predictive models [20]. Temperature models that incorporate wind, such as those proposed by Skoplaki, Mattei and Kurtz, elucidate the implications of wind speed for the thermal performance of PV modules, which can mitigate overheating and enhance energy conversion efficiency [33]. The absence of wind modelling in the majority of PV simulations results in less accurate predictions of actual performance patterns, especially in regions characterised by high variability in wind conditions, such as South Africa.

Furthermore, many simulation models tend to oversimplify environmental assumptions and fail to account for regional climatic variations, such as dust accumulation, which is a notable characteristic of the Free State region, and its subsequent impact on PV system performance [11]. The lack of granular data contributes to the discrepancy between model predictions and actual performance, which is often influenced by a complex interplay of factors that are challenging to replicate in conventional models. In the case of dust accumulation, which significantly diminishes the performance ratio of PV systems by reducing the solar irradiance incident on the panels, this issue is particularly relevant to the dry and arid regions of South Africa [3].

This study addressed these research gaps by incorporating real-world operational data from the CUT's dual-axis PV system, a notable case study situated in Bloemfontein, South Africa. The study utilised data from one tracker as a benchmark, processing it to obtain aggregated values for the entire solar plant. By leveraging these data, the model can more accurately reflect actual environmental conditions and system performance. Additionally, this research incorporated wind-related temperature models (e.g., [145]) to better account for the effects of wind speed, which have been demonstrated to reduce module temperature and, consequently, thermal losses, thereby enhancing system efficiency. The application of ML techniques, such as SVMs and ensemble trees, allowed the model to capture the complex, non-linear relationships between environmental variables and PV output to provide a more robust and flexible tool for performance prediction [13]. These advancements are essential for ensuring that predictive models for dual-axis PV systems more accurately reflect real-world conditions and, by extension, improve the planning, design, and operation of such systems in South Africa and similar climates.

In conclusion, the insights derived from this study not only bridge the gap between simulations and real-world operations but also provide a more reliable tool for future PV plant planning in South Africa. The enhanced model, which incorporated environmental factors such as wind speed, temperature effects, and local climate variability, offers valuable contributions to the optimisation of PV systems in regions characterised by harsh environmental conditions. This research aimed to advance the field of PV system modelling by contributing to more accurate energy yield forecasting and the sustainable development of renewable energy in South Africa.

CHAPTER 3:
MODELLING AND VALIDATION

3.1 INTRODUCTION

This chapter delineates the steps and methodologies employed in the development and validation of a predictive model for dual-axis PV tracking systems. An average tracker unit was utilised, and the data were interpolated to derive a global aggregate representative of the entire plant. The focus was on creating an accurate simulation of the operational behaviour of PV systems through ML methods, with the aim of predicting energy outputs based on actual field scenarios. The model aspired to encompass a wide array of environmental variables, including solar irradiance, ambient temperature, wind speed, and humidity, which influence the efficacy of PV systems under dynamic outdoor conditions. By leveraging data from the CUT's PV power station in South Africa, this study ensured that the model was grounded in realistic, locally relevant scenarios in order to enhance the understanding of the behaviour of dual-axis tracking systems within specific regional contexts.

Model validation was a crucial aspect of this approach, which ensured the reliability and accuracy of the model in practical applications. The datasets obtained from the CUT's PV power plant comprised detailed environmental and operational parameters that served as the foundation for model training and validation. ML models adept at managing complex non-linear relationships among variables were employed in the model formulation. Furthermore, the model underwent rigorous testing and evaluation through the comparison of predicted outcomes with the actual energy output, which affirmed its robustness and accuracy. This approach guaranteed that the model was not only scientifically defensible but also practically useful in predicting energy generation and optimising operational strategies for PV systems within the study region.

3.2 DATA COLLECTION AND PREPROCESSING

Data acquisition is a critical phase of the modelling process in PV systems due to its direct impact on the accuracy and efficacy of predictive models. In this study, data were collected at the CUT's PV site located in Bloemfontein, South Africa. Rather than measuring every individual unit, data from a single standard tracker were employed and extrapolated using interpolation to simulate the entire system. The site features a dual-axis PV tracking system that provides a valuable real-world dataset for model training. Meteorological stations and the Sunny Portal monitoring system served as the primary data sources, which played a fundamental role in the collection of various environmental and operational parameters. Figures 3.1 and 3.2 illustrate the CUT's dual-axis PV tracking system and the weather station from which climate condition data were obtained, as sourced from the Southern African Universities Radiometric Network (SAURAN) instruments stationed at the CUT (Latitude: -29.121337; Longitude: 26.215909; Elevation: 1 397 m) in Bloemfontein.

Figure 3.1: Dual-axis PV tracking systems installed on the Central University of Technology, Free State's (CUT) premises [117]

Figure 3.2: Weather station (Southern African Universities Radiometric Network [SAURAN]) on a CUT rooftop [117]

3.2.1 Environmental variables

The power and energy output of a PV system is highly dependent on site-specific environmental parameters. Precise quantification of these parameters is therefore essential for any predictive study of the PV system. Real-time environmental data for the study was sourced from the PV system at the CUT, encompassing a range of the most influential known determinants of PV output.

One of the most critical variables is GHI, which measures the total solar radiation incident on a horizontal surface and serves as a primary indicator of solar resource availability at a particular location [15]. DNI was also measured, since it represents the component of solar radiation that comes directly from the sun in a straight line [118]. DNI is particularly relevant for dual-axis tracking systems that adjust the angle of the panels to optimise direct sun exposure [118]. Conversely, DHI accounts for solar radiation that is scattered by molecules and particles in the atmosphere, thus aiding in the estimation of indirect sun exposure [119].

Ambient temperature constitutes another key environmental parameter that influences PV cell performance. As cell temperature rises, PV efficiency typically declines due to increased internal resistance and diminished voltage output [20]. This relationship highlights the necessity of measuring not only irradiance but also the effects of temperature. Wind speed and wind direction are similarly significant; wind can cool PV modules, which reduces their operational temperature and indirectly enhances energy conversion efficiency [33]. Additionally, wind direction can provide insight into how airborne dust might accumulate on panel surfaces, which can potentially result in soiling-related losses [120].

System efficiency is also affected by relative humidity in semi-arid or coastal climates. Elevated humidity levels can lead to the deposition or condensation of moisture on panel surfaces, which results in increased light scattering and a reduction in effective solar irradiance impacting the cells [121]. Energy yield is optimised when the tilt angle of the PV module and the azimuth angle are carefully calibrated. The tilt angle determines the angle at which solar radiation is received throughout the seasons, while the azimuth refers to the compass direction of the panel faces, both of which are critical for ensuring stable sun exposure at all times of the day [53].

Seasonal factors also play a role; as the sun's angle shifts with the seasons, the angle and intensity of incoming solar radiation change accordingly. Aggregating irradiance data throughout the day allows the model to capture diurnal variation in solar energy availability [122]. Ultimately, the actual PV power output measured through the Sunny Portal platform is utilised for the validation and refinement of the predictive model. Sunny Portal provides high-resolution operational data, encompassing power generation, voltage, and temperature parameters at both the system and inverter levels [123], which enable the model output to be compared against actual performance parameters.

3.2.2 Data preprocessing

To ensure high-quality inputs for model development, the obtained data underwent a rigorous preprocessing phase. Such a protocol is requisite in all ML applications, as the accuracy and efficacy of the predictive model significantly depend on the cleanliness, organisation, and normalisation of the training set [124],[125]. The system performance and environmental information, initially collected from a variety of sources, including the Sunny Portal online network and internal meteorological instruments, were first compiled. The data were stored in Microsoft Excel format, and the xlsread() function of MATLAB was employed to import the well-structured variables into the computing environment with high fidelity and column-wise integrity [126].

A comprehensive data-cleaning protocol followed extraction. During this stage, missing, inconsistent, or anomalous values were identified and addressed. Missing data points – often the result of sensor faults or transmission errors – were rectified through interpolation methods such as the interp1() linear interpolation function of MATLAB. When data could not be recovered or were significantly erroneous, they were excluded to prevent the introduction of bias or variance into the model [127]. These cleaning processes are particularly critical in time series applications, where a small number of corrupted timestamps can adversely affect model performance [128].

Subsequently, each of the continuous numeric variables was scaled using min-max scaling to the interval (0, 1). Normalisation aids in treating the variables equally, which ensures that variables with larger magnitudes, such as temperature or irradiance, do not disproportionately influence those with smaller scales. This facilitates the rapid convergence of models such as SVMs and ANNs [129]–[131].

Given the time-varying nature of PV power generation, the dataset was structured into a chronological time series format. This was accomplished using the table data type in

MATLAB, which supports multi-variable time ordering, rapid querying, and seamless integration with the .mat file format utilised in model development [23]. The timestamps were standardised and synchronised across all features to preserve the temporal integrity of the data, which is essential for dynamic forecasting applications [124].

Finally, a feature engineering and selection phase was conducted. Redundant or unnecessary variables were eliminated using correlation analysis and principal component methods. Conversely, new synthetic variables were created to enhance model learning. These included combinations of solar radiation components such as GHI, DNI, and DHI to ascertain total irradiance. Additionally, interactive weather effects such as temperature-humidity indices or wind-chill factors were calculated to account for the cumulative effects on PV output [130]. These artificially created variables provided more refined inputs, which enhanced the model's capacity to learn intricate environmental dependencies.

3.2.3 Model training and validation

The preprocessed and well-structured data were subsequently inputted into MATLAB's Regression Learner App to assess a variety of ML algorithms and evaluate their performance. Training incorporated cross-validation methods, such as five-fold cross-validation, to prevent overfitting and ensure the robustness of the model. The models were evaluated against principal performance indicators, including the RMSE, MAE, and R^2, which are measures of prediction accuracy and the ability to generalise.

Through the implementation of this comprehensive preprocessing pipeline, the data were successfully transformed into a format that was conducive to the development of high-accuracy predictive models for PV system performance.

3.2.4 Diagram and visual representation

Figure 3.3 illustrates how the data preprocessing pipeline was structured in MATLAB:

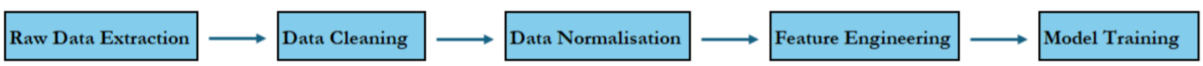

Figure 3.3: Data preprocessing flowchart

3.3 MODEL DEVELOPMENT

3.3.1 Methodology for developing a prediction model for Photovoltaic power output

The primary objective of this study was to address the research gap by developing and validating an extremely accurate and dependable predictive model for the energy output and performance of dual-axis PV tracking systems in South Africa, with specific application to the CUT installation in Bloemfontein. The model was designed to forecast energy production over time by utilising environmental parameters in order to provide insight into the future performance of PV systems under varying conditions. The MATLAB Regression Learner App was employed, leveraging ML algorithms for the design and validation of the model [23].

3.3.1.1 Step 1: Preparing and importing data

The initial step involved the importation of environmental and performance data, specifically GHI, DNI, DHI, ambient temperature, wind speed, and relative humidity. These data were obtained from the Sunny Portal system, which provided real-time energy output information from the CUT's dual-axis PV tracking system. The data were exported in MATLAB format from a Microsoft Excel spreadsheet that contained the daily environmental variables.

The data were structured into a tabular format, aligning the environmental predictor variables with the response variable (power output of PV). This organised dataset served as the foundation for model development [3].

3.3.1.2 Step 2: Training the model

Subsequently, the model was trained using the Regression Learner App in MATLAB. This software allowed the researcher to experiment with a variety of ML algorithms, including decision trees, linear regression, SVMs, ensemble methods, and Gaussian process regression. All models were evaluated against performance metrics such as RMSE, R^2, and MAE. The steps in the training process included:

1. Data loading: The input features and the target variable, specifically the power output of the PV system, were loaded into the software.

2. Model selection and testing: A variety of algorithms were evaluated to identify the one that exhibited the best performance. Key metrics, including RMSE, R^2, and MAE, were employed to assess the models and to select the algorithm characterised by a low RMSE and a high R^2 [14]. Table 3.1 presents the multiple algorithms that were tested.

Table 3.1: The algorithms that were tested

Favourite	Model no.	Model type	Status	RMSE (validation)	MSE (validation)	R^2 (validation)	MAE (validation)
0	1	Tree	Trained	0.4126	0.17024	0.99277	0.16616
0	2.1	Linear regression	Trained	1.3958	1.9483	0.91725	1.0325
0	2.2	Linear regression	Trained	1.0047	1.0094	0.95713	0.66731
0	2.3	Linear regression	Trained	1.4571	2.1232	0.90982	1.0086
0	2.4	Stepwise linear regression	Trained	0.93453	0.87334	0.96291	0.66698
0	2.5	Tree	Trained	0.4126	0.17024	0.99277	0.16616
0	2.6	Tree	Trained	0.47568	0.22627	0.99039	0.21281
0	2.7	Tree	Trained	0.62186	0.38671	0.98358	0.29604
0	2.8	SVM	Trained	1.3872	1.9242	0.91827	1.0226
0	2.9	SVM	Trained	1.0195	1.0394	0.95585	0.675
0	2.10	SVM	Trained	11.865	140.77	-4.9787	1.2727
0	2.11	SVM	Trained	0.89253	0.7966	0.96617	0.62644
0	2.12	SVM	Trained	0.76489	0.58505	0.97515	0.52961
0	2.13	SVM	Trained	1.0712	1.1474	0.95127	0.768
0	2.14	Efficient linear	Trained	1.5217	2.3156	0.90165	1.2076
0	2.15	Efficient linear	Trained	1.5097	2.2792	0.9032	1.162
0	2.16	Ensemble	Trained	0.48744	0.2376	0.98991	0.30715
1	2.17	Ensemble	Trained	0.39896	0.15917	0.99324	0.16795
0	2.18	Gaussian process regression	Trained	0.65441	0.42826	0.98181	0.30576
0	2.19	Gaussian process regression	Trained	0.65151	0.42446	0.98197	0.28159

Favourite	Model no.	Model type	Status	RMSE (validation)	MSE (validation)	R^2 (validation)	MAE (validation)
0	2.20	Gaussian process regression	Trained	0.50426	0.25428	0.9892	0.23098
0	2.21	Gaussian process regression	Trained	0.55613	0.30928	0.98686	0.23998
0	2.22	Neural network	Trained	0.73689	0.54301	0.97694	0.35694
0	2.23	Neural network	Trained	2.0186	4.0747	0.82694	0.36227
0	2.24	Neural network	Trained	1.2817	1.6428	0.93023	0.3269
0	2.25	Neural network	Trained	1.1471	1.3158	0.94412	0.31094
0	2.26	Neural network	Trained	0.5614	0.31517	0.98661	0.26201
0	2.27	Kernel	Trained	1.5374	2.3637	0.89961	0.99431
0	2.28	Kernel	Trained	1.1589	1.3431	0.94296	0.73306

The best model, identified as the ensemble tree model, showed superior performance metrics:

- RMSE: 0.39896 (lowest RMSE)
- R^2: 0.99324 (highest R^2)
- MAE: 0.16795 (lowest MAE)

This model was selected for its exceptional capacity to generalise across various environmental conditions and timeframes, which rendered it suitable for long-term predictions.

3.3.1.3 Step 3: Predicting with the trained model

Upon completion of the training and validation phases, the model was prepared for application in generating predictions for novel data. Predictions for PV power output were derived by inputting new data into the trained model. The new input data were formatted in a manner consistent with the training data to ensure uniformity in the variables employed. The prediction process adhered to the steps outlined in Figure 3.4.

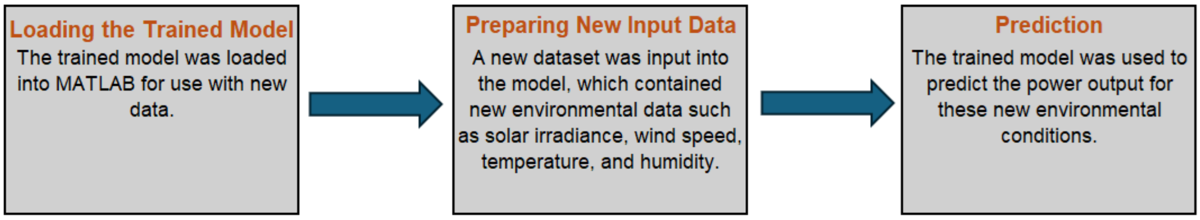

Figure 3.4: Steps process of prediction

3.3.1.4 Step 4: Visualising and saving predictions

To assess the performance of the model, a comparative analysis was conducted between the predicted and actual values. Graphical representations of the predicted values were generated using the plotting functions available in MATLAB. The actual and predicted values were juxtaposed by plotting them on the same graph, as illustrated in Figure 3.5, which depicts the plotted curves of both actual and predicted values.

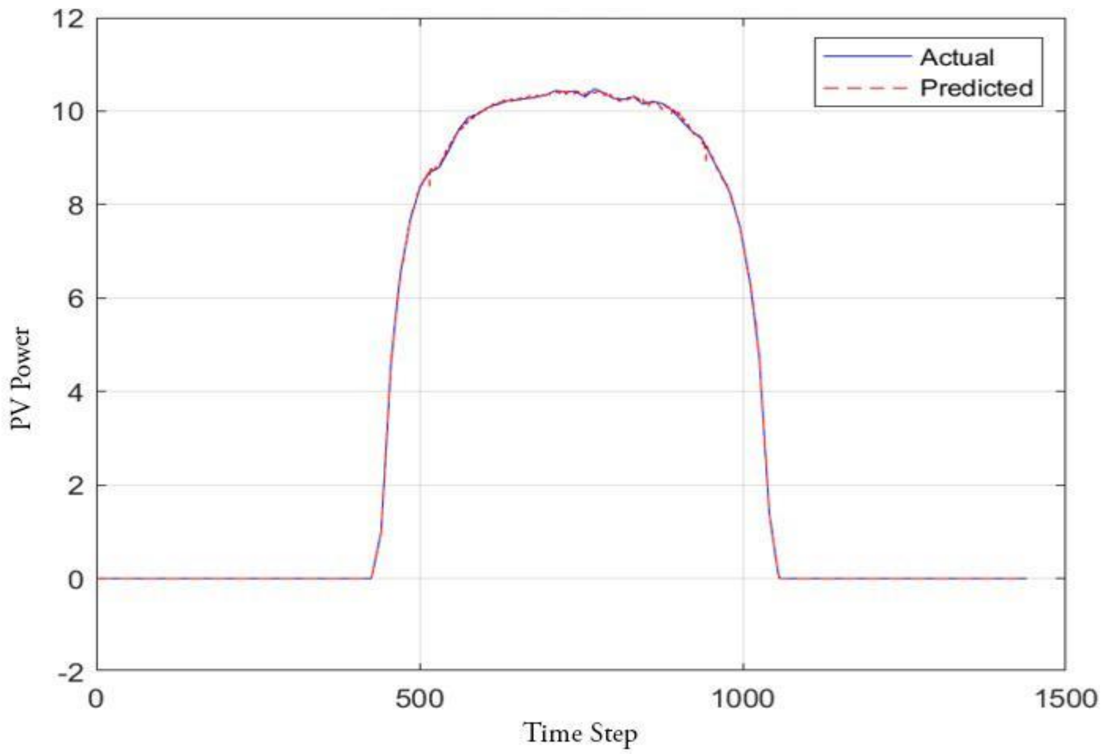

Figure 3.5: Plotted curve of predicted and actual values

3.3.1.5 Step 5: Model evaluation and performance metrics

The performance of the model was evaluated using various metrics. These metrics were derived from the training and validation datasets to verify the robustness of the model. The following key performance indicators were utilised [3]:

- RMSE: Measured the model's error by comparing predicted and actual values.
- MAE: Measured the average magnitude of errors in a set of predictions.
- R^2: Indicated how well the model's predictions approximated the real data.

3.3.2 Conclusion and diagram support

The overall strategy for developing and training the models was robust and characterised by a structured approach encompassing data preprocessing, feature set selection, model training, and testing. Although no diagrams were utilised in this segment, it would be beneficial to present flowcharts or diagrams that illustrate the data pipeline and model structure to enhance clarity. Additionally, plots depicting performance metrics (e.g., "Predicted vs actual PV power output") and graphs illustrating model error rates could be provided for visual inspection.

3.4 MODEL VALIDATION AND PERFORMANCE METRICS

To ensure the reliability and generalisability of the predictive model, k-fold cross-validation was employed. In this method, the training data were divided into k equally sized subsets, and the model was iteratively trained on k-1 folds while being tested on the remaining fold. This validation process was repeated for k iterations, and the average performance metrics were computed to mitigate the effects of dataset bias and overfitting [24]. A 10-fold cross-validation approach was implemented in this study.

The calculated PV power output levels were compared with actual measurements obtained from the Sunny Portal system. To assess the quantitative performance of the model, three prominent statistical indicators were employed.

In the experiment, the ensemble tree model achieved a very high R^2 value of 0.99324, RMSE of 0.39896, and MAE of 0.28715, which indicated high agreement between predicted and actual PV output levels with small average deviations.

These metrics led to the conclusion that the model effectively captured the non-linear interactions between environmental variables (such as solar irradiance, temperature, and relative humidity) and power production with a high degree of precision.

These results affirmed that the PV power forecasting model, utilising a combination of sensor and weather data, was a reliable predictive tool that can support operational decisions pertaining to solar power systems.

3.5 CONCLUSION

Despite the predictive model demonstrating strong performance, challenges and limitations emerged during the modelling and testing phases that may impact its scalability and robustness in real-world deployment scenarios.

A primary issue was the incompleteness of the data, which arose from interruptions in data recording due to hardware failures, network connectivity issues, or power supply loss. Such gaps in the data hindered the model's ability to learn temporal patterns and reduced the accuracy of forecasts across all test scenarios. In particular, gaps occurring during periods of high variability (e.g., cloudy mornings or abrupt changes in irradiance rates) introduced noise and limited the comprehensiveness of the training dataset [25].

Another challenge involved sensor calibration discrepancies, particularly concerning temperature and irradiance measurements. Erroneous or inconsistent sensor readings, resulting from ageing, improper calibration, or contamination, led to deviations in the quality of the input data. Given the model's high dependency on environmental parameters, these deviations could propagate through the learning process and diminish the quality of the output. Systematic errors in sensors can severely compromise the performance of data-driven energy forecasting models [26].

Furthermore, the model's limitations included the impact of environmental variability factors such as dust accumulation, shading from vegetation or other structures, and degradation due to environmental conditions. These factors are both location-dependent and time-varying, yet could not be directly incorporated into the model due to the absence of live monitoring or correction of the variability parameters. Consequently, the model was likely to lose precision in applications where such variability leads to significant deviations from historical behaviours [28].

These challenges underscore the need for robust data preprocessing, regular maintenance of the sensor module, and the application of adaptive modelling methods that account for environmental anomalies. Future improvements should involve the integration of external satellite data, automatic anomaly detection, and the development of hybrid models that combine physics-based insights with ML techniques.

CHAPTER 4:
RESULTS AND DISCUSSIONS

4.1 INTRODUCTION

This chapter reports and interprets the results from the predictive model formulated in Chapter 3, specifically in terms of accurately estimating the energy production of a dual-axis PV tracking system. The performance of the model is assessed using a number of statistical indicators, namely RMSE, MAE, and R^2. These parameters facilitate a rigorous estimation of the model's precision and its generalisability in a realistic environment. Furthermore, plots comparing estimated and actual PV production were presented to verify the efficacy of the model under various operational scenarios and environmental conditions. The chapter also presents the results of the study concerning the impact of the primary environmental parameters – solar irradiance, ambient temperature, wind speed, and relative humidity – on both the model's precision and the actual production of the system.

In addition to performance evaluation, this chapter addresses long-term trends through a degradation analysis of the CUT's dual-axis PV system over a five-year period (2019 to 2023). This analysis revealed a gradual decline in energy output due to factors such as soiling, component ageing, and environmental stressors, which are critical aspects for understanding the operational lifespan and maintenance needs of PV systems. The results were employed to inform projections of energy yield over the next two decades to provide a forward-looking perspective on the system's potential and limitations. Long-term forecasting and degradation modelling are essential tools for energy planners and stakeholders, as they assist in optimising maintenance strategies and investment decisions [3]–[29]. Overall, this chapter offers valuable insight into the dynamic performance of dual-axis PV systems in semi-arid regions and underscores the importance of predictive analytics in renewable energy infrastructure planning.

4.2 MODEL PERFORMANCE EVALUATION

To thoroughly assess the performance of the predictive model, a series of key statistical indicators were analysed, namely RMSE, MAE, and R^2. These indicators elucidate the magnitude and distribution of prediction errors and the model's ability to replicate observed patterns of power production.

The ensemble tree model was identified as the best-performing model following rigorous testing across various ML methods. Table 4.1 provides a summary of the results of the tested models using the Regression Learner App in MATLAB. It was observed that the ensemble tree model yielded an RMSE of 0.39896 and an R^2 of 0.99324, which outperformed other alternatives such as linear regression, SVMs, and neural networks.

Table 4.1: Comparative performance of regression models

Model type	RMSE	MAE	R^2
Ensemble tree	0.39896	0.16795	0.99324
Linear regression	1.3958	1.0325	0.91725
Neural network	0.5614	0.26201	0.98661
SVM	1.0712	0.768	0.95127

The R^2 value of 0.99324 indicated that the model accounted for up to 99.32% of the variability in actual power output, thus demonstrating a highly satisfactory fit to the data. A significant advantage of PV forecasting systems lies in the precision of their predictions, which enhances the reliability of operational planning [29].

The RMSE is defined as the square root of the average of the squares of the differences between actual and predicted values. The low RMSE of 0.39896 confirmed that substantial deviations between predicted and actual values were minimal.

Similarly, the MAE, calculated as the average of the absolute differences between actual and predicted values, was approximately 0.16795. This reflects the model's accuracy across individual predictions. Unlike the RMSE, the MAE provides a more intuitive indication of average prediction error without disproportionately penalising large errors [30]. Figure 4.1 illustrates the overlay of predicted and actual values using time series visualisation. The close alignment of the curves visually validates the numerical metrics.

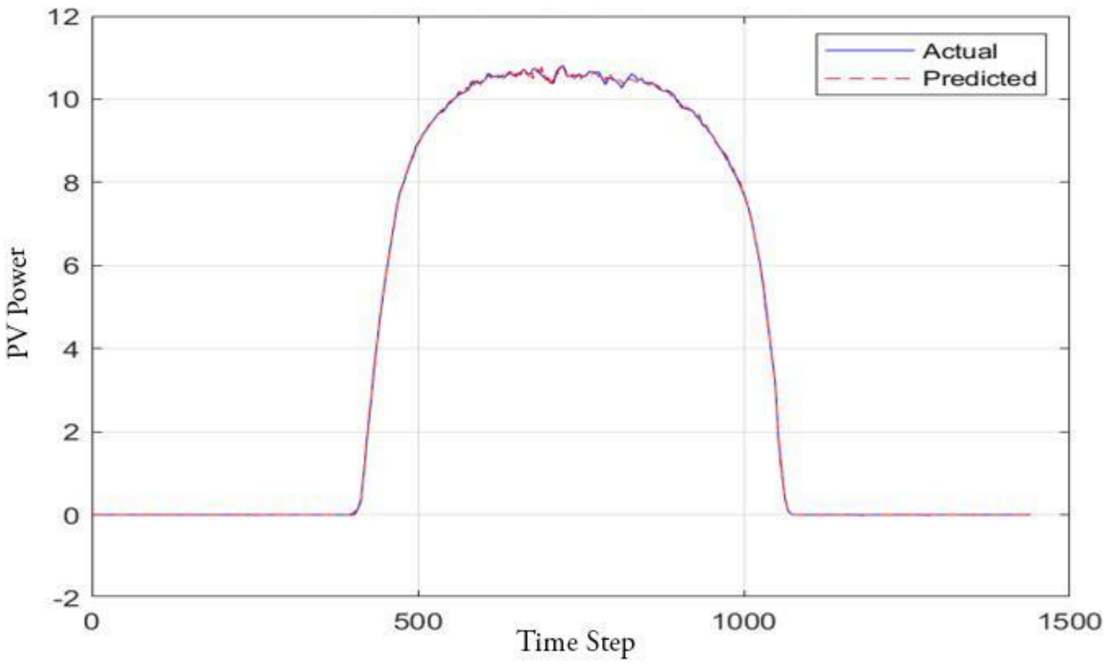

Figure 4.1: Predicted versus actual PV output over a day (22 December 2020)

Apart from statistical precision, ensemble trees are highly suitable for solar data as they are capable of representing complex non-linear relationships and handling multicollinearity among the input variables [14]. Solar irradiance, temperature, wind speed, and relative humidity exhibit complex non-linear relationships that cannot be adequately represented through conventional regression models.

The strength of the ensemble tree model lies in its ability to average decision trees using methods such as bagging or boosting. This ensemble strategy helped to minimise variance and stabilise the model [16]. More importantly, the interpretability of the tree structure elucidated the importance of various variables, which was of great significance for understanding the environmental factors that affect PV output.

Figure 4.2 presents a bar chart that illustrates the relative importance of each predictor variable in the ensemble tree model. DNI, GHI, and ambient temperature ranked at the top in terms of influence, thereby confirming previous empirical findings.

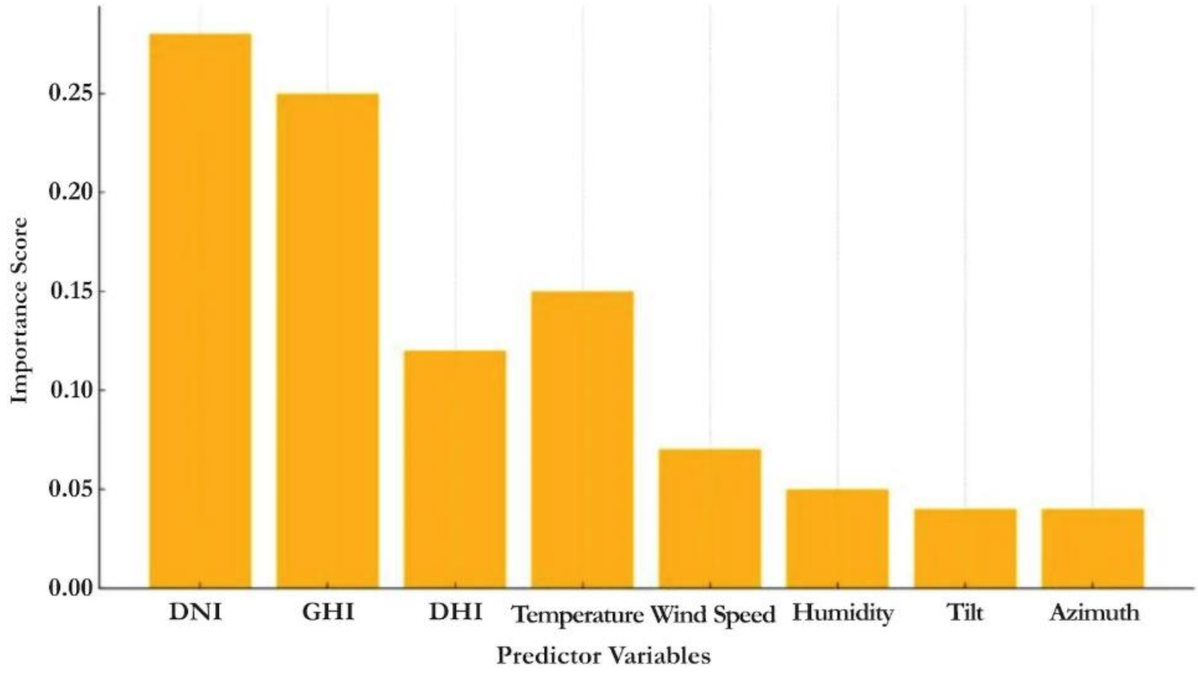

Figure 4.2: Predictor importance in the ensemble tree model

Additionally, the robustness of the model was assessed by employing it on unseen validation data, which comprised various days with differing weather conditions. The results consistently demonstrated reliability, with validation R^2 values exceeding 0.98 for all tested days. This finding supports the model's utility in real-world conditions and its potential for integration into operational forecasting tools for solar farms.

In summary, the ensemble tree model exhibited clear superiority in estimating PV power output, characterised by strong precision, generalisability, and interpretability. These attributes render it highly suitable for supporting decision-making processes in system maintenance, energy trading, and long-term energy planning.

4.3 COMPARATIVE ANALYSIS WITH EXISTING MODELS

Traditional PV simulation software, such as PVsyst, RETScreen, and SAM, is used in both academic and practical domains for simulating the performance of PV systems. These software tools employ physical and empirical models to forecast energy yield under normal or typical meteorological year conditions. However, they frequently fail to respond dynamically to real-time environmental changes, such as sudden variations in irradiance, dust accumulation, or short-term weather irregularities [31].

For instance, PVsyst is proficient in simulating fixed-tilt and tracking systems under controlled inputs, yet it relies heavily on predefined datasets and does not incorporate real-time feedback into the simulation [32]. RETScreen is more focused on feasibility studies and techno-economic analyses than on day-level forecasting, which limits its applicability in operational decision support. SAM provides detailed performance simulations but lacks the integration of artificial intelligence (AI) and ML elements necessary for calibrating forecasts in real time using sensor inputs.

In contrast, the ML model presented in this study demonstrated significant strengths in flexibility, reactivity, and accuracy. By incorporating the ensemble tree algorithm through the Regression Learner App in MATLAB and utilising meteorological inputs sourced from the CUT's PV plant, the model was responsive to field variability that conventional software cannot adequately track. Table 4.2 outlines the comparison between conventional tools and the development of the ML model [54]–[69],[124]–[130].

Table 4.2: Comparison between traditional tools and the developed ML model

Feature	**PVsyst**	**RETScreen**	**SAM**	**Developed ML model**
Real-time weather input	No	Limited	No	Yes
AI/ML algorithm support	No	No	No	Yes (ensemble tree)
Sensor feedback integration	No	No	Partial	Yes
Environmental sensitivity	Moderate	Low	Moderate	High
Adaptability to new conditions	Low	Low	Medium	High
Accuracy in short-term forecasting	Medium	Low	Medium	High

Furthermore, the developed model demonstrated enhanced predictive realism through the integration of the thermal models of wind, as presented in Skoplaki et al. [145] and Mattei et al. [33]. These models indicate that the operating temperature of PV panels is significantly influenced by wind, which in turn directly affects efficiency. Standard simulation models typically utilise fixed derating factors or simplistic thermal models that fail to consider spatial cooling effects [33].

The inclusion of time-varying environmental parameters, such as instantaneous wind direction and humidity, enables the ensemble tree model to adapt more effectively to environmental changes. For instance, during stormy or dusty conditions, the model dynamically adjusts its output predictions, contrasting with the static or average conditions assumed by conventional models. This adaptability is particularly advantageous in semi-arid regional climates, such as that of the Free State, where environmental parameters can fluctuate significantly over short periods. Figure 4.3 illustrates the model's flexibility by comparing the instantaneous ML model prediction curve with the static output profile of PVsyst on a selected day that is characterised by high variability.

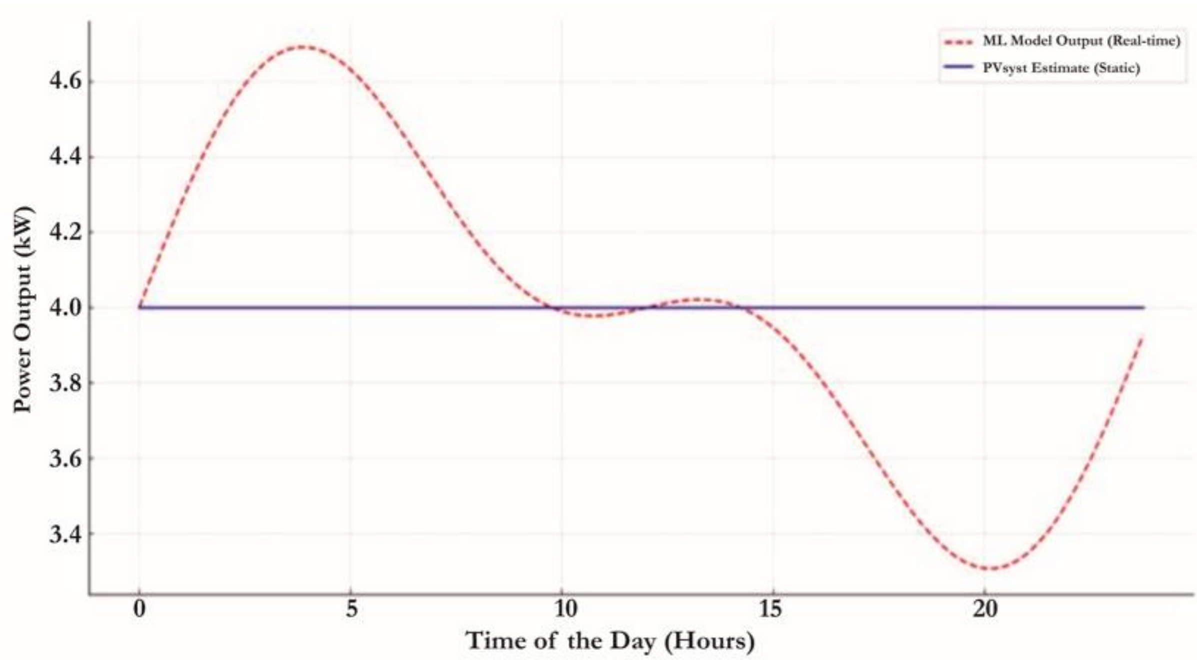

Figure 4.3: Real-time ML model versus PVsyst output (variable weather conditions)

The strength of the developed model lies not only in its scalability but also in its capacity for continual improvement through ongoing learning. As operational data accumulate, the model can be easily retrained or fine-tuned, which enhances its precision over time. In contrast to conventional tools that necessitate manual updates and often fail to account for older equipment, seasonal degradation, or unforeseen system behaviours, this model offers a more adaptive solution.

In conclusion, the comparative study elucidated the limitations of fixed simulation systems while highlighting the advantages of dynamic, data-driven systems. The proposed ensemble tree-based model exhibited superior flexibility, accuracy, and environmental consciousness, which established it as a valuable instrument for the online monitoring and decision support of PV systems.

4.4 PRACTICAL IMPLICATIONS

The development of a reliable ML predictive model for PV systems presents numerous practical applications that extend beyond theoretical performance assessments. This section delineates the model's utilisation in operational decision making, predictive maintenance, energy forecasting, and strategic planning, which can enhance overall PV system efficiency and reliability.

4.4.1 Predictive maintenance scheduling

One of the primary applications of the model is in predictive maintenance. By continuously forecasting and monitoring PV generation output, operators can identify potential performance degradation linked to environmental factors such as excessive dust accumulation, high humidity, or irregular temperature variations. A significant deviation between predicted and actual output may indicate issues such as shading, inverter malfunctions, or soiling. This predictive capability facilitates condition-based maintenance scheduling, as opposed to traditional periodic inspections, which prevents unnecessary downtime and optimises plant operation [34].

Mathematically, predictive maintenance can be bolstered by analysing the residual error, where peaks may signify anomalous events that require attention. Figure 4.4 illustrates the rolling mean of the residuals of errors over a week.

While the legend in Figure 4.4 includes a marker designated for maintenance flags, no such flags are evident in the plot, as the residual errors remained well within the specified operational range (typically ±0.5 kW). This suggests that the PV system operated within normal performance parameters during the monitored period.

The residual monitoring approach is commonly employed in PV performance assessments, where error thresholds are utilised to track discrepancies between actual and predicted outputs [131],[132]. The absence of maintenance flags in the current analysis served as a clear indicator of stable system operation, with no significant performance anomalies warranting maintenance intervention.

Threshold fault detection methods are particularly effective in online monitoring scenarios, as they strike a balance between sensitivity and false alarms [15]–[133]. The model's alignment with predicted outputs indicated that the PV system functioned reliably under the environmental and operational conditions recorded during the specified timeframe.

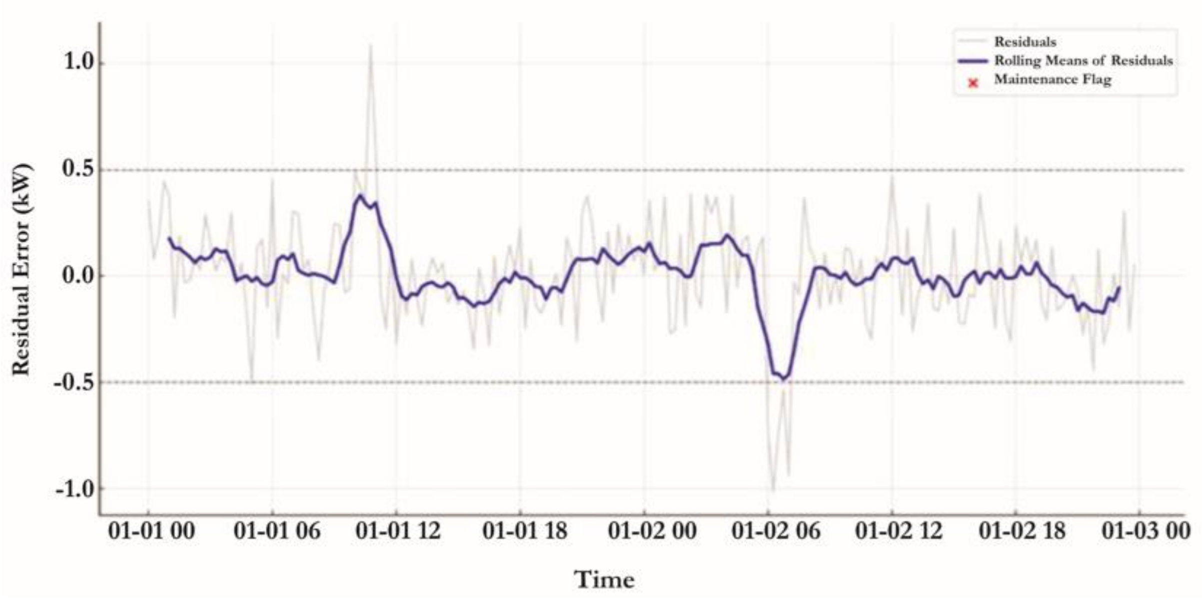

Figure 4.4: Rolling mean of residual errors for fault detection

4.4.2 Improved grid integration

Accurate PV output forecasting is essential for the integration of solar power into the electrical grid, particularly in regions such as South Africa, where the contribution of solar energy is becoming increasingly significant to the energy mix. Unpredictable fluctuations in PV output may induce instability in the grid. The high-resolution forecasts generated by the model empower utility operators to make informed strategic decisions regarding load balancing, reserve capacity, and dynamic demand-response mechanisms [36]. Table 4.3 provides a summary of the comparison between the average lead time and the accuracy of the conventional models and the proposed ensemble tree models.

Table 4.3: Forecasting accuracy and grid readiness

Model type	Avg. forecast lead time	Accuracy (R^2)	Grid-friendliness
PVsyst	1 day	~0.91	Moderate
Neural network	1 hour	~0.98	High
Ensemble tree model	15 minutes	0.993	Very High

4.4.3 Enhanced energy yield forecasting

Accurate short- and long-term forecasts of energy yield are crucial for stakeholders, including investors, policymakers, and system operators. The ensemble tree model effectively balances environmental variability with more reliable forecasts in comparison to fixed-parameter models. Consequently, it facilitates more realistic planning of energy production, the design of tariff rates, and the establishment of performance warranties.

Furthermore, cumulative yield estimation plays a significant role in identifying seasonal trends and assessing the economic viability of expansion plans. Figure 4.5 illustrates the monthly estimated and actual yield over a one-year period, which reveals a high correlation and seasonal performance data.

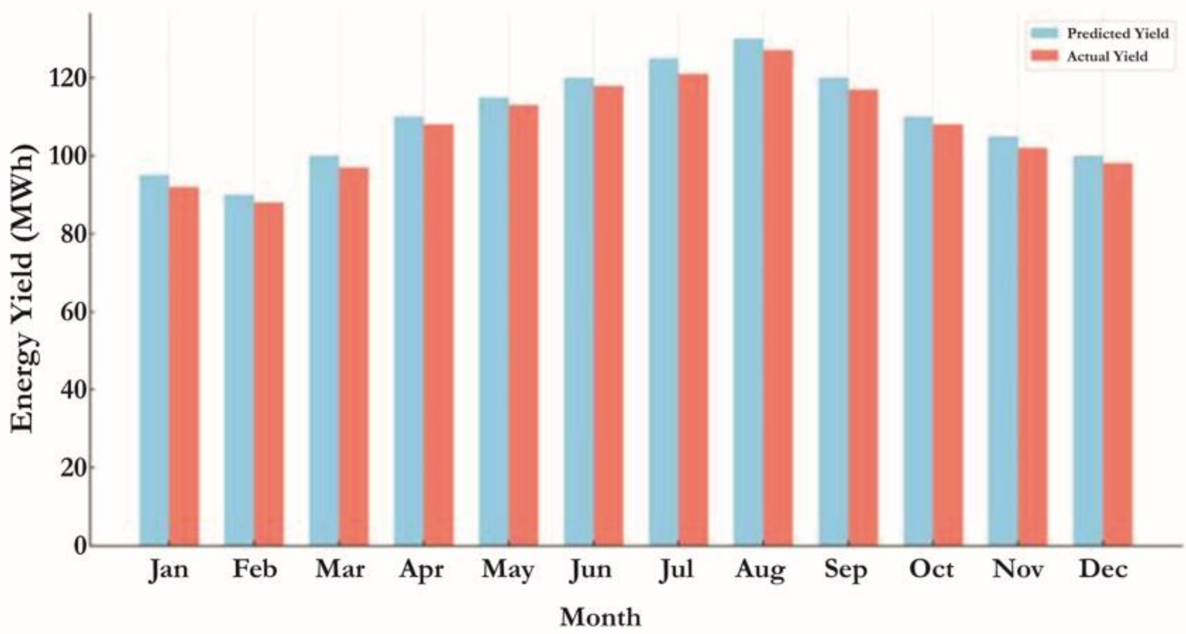

Figure 4.5: **Monthly predicted versus actual energy yield**

4.4.4 Decision support for system upgrades and expansion

The ability to forecast the impact of system modifications or environmental changes on PV generation enhances informed decision making regarding system upgrades. For example, the model can simulate the effects of integrating new tracking mechanisms or installing cleaning robots by adjusting the primary input variables. Subsequently, scenario analysis is performed to assess system design optimisation and ROI [36].

A decision matrix, as illustrated in Table 4.4, is applicable for ranking investment options based on the outcomes generated by the model.

Table 4.4: Scenario-based decision support matrix

Upgrade scenario	**Cost impact**	**Yield improvement**	**ROI (3 years)**	**Recommended?**
Robotic cleaning system	Medium	High	18%	Yes
Fixed to dual-axis upgrade	High	Very high	25%	Yes
Improved inverter	Low	Moderate	12%	Yes
Anti-soiling coating	Medium	Low	5%	No

These actionable insights empower PV operators and planners to make evidence-based decisions that enhance system efficiency and ROI.

4.5 DISCUSSION OF THE FINDINGS

The results of the study revealed the high precision and robustness of the ensemble tree model in forecasting PV power generation under various environmental scenarios. Notably, the validation of the model demonstrated a close correspondence between the predicted and actual energy yields, which supports the applicability of ensemble-related ML models in the task of solar power forecasting. This is evidenced by the excellent R^2 value (0.99324) and low RMSE (0.39896), which validated that the model adequately captured the non-linear interactions between various environmental inputs and PV performance.

A significant aspect of this investigation was the degradation test conducted on the CUT's PV system over the period from 2019 to 2023. The test indicated a cumulative performance loss of approximately 3.96%, which was consistent with degradation rates documented in the literature for comparable systems operating in semi-arid climates [3]. Such a gradual decline in performance results from a combination of environmental and operational factors. Among these factors, dust accumulation on the panel surfaces (soiling) is the most prominent, as it reduces exposure to irradiance. Mani and Pillai report that soiling losses in dusty regions can decrease energy yield by up to 20% in the absence of regular cleaning of the panels [22].

Another contributing factor to performance loss is ageing equipment. The materials in PV modules, inverters, and tracking systems gradually degrade with prolonged exposure to UV radiation, thermal cycling, and humidity effects. This condition is typically referred to as PV module fatigue and results in a gradual decline in module efficiency and electricity conversion rates [37]. Table 4.5 summarises typical causes of degradation and the estimated loss in system performance[22],[134]–[135].

Table 4.5: Common PV degradation factors and estimated impact

Cause	**Impact range (%)**
Dust/soiling	1–20
Humidity and moisture ingress	0.5–3
Temperature cycling	0.5–2
Ageing of inverters	0.5–1

In addition to long-term deterioration, short-term environmental variability also exerts a significant influence on system performance. For instance, wind speed has a cooling effect on PV modules, which in turn reduces cell temperature and enhances electrical efficiency. Conversely, elevated humidity levels may diminish the transparency of the panels or contribute to surface condensation, which can slightly reduce solar transmittance. Figure 4.6 illustrates the inverse relationship observed between panel temperature and wind speed on a selected summer day.

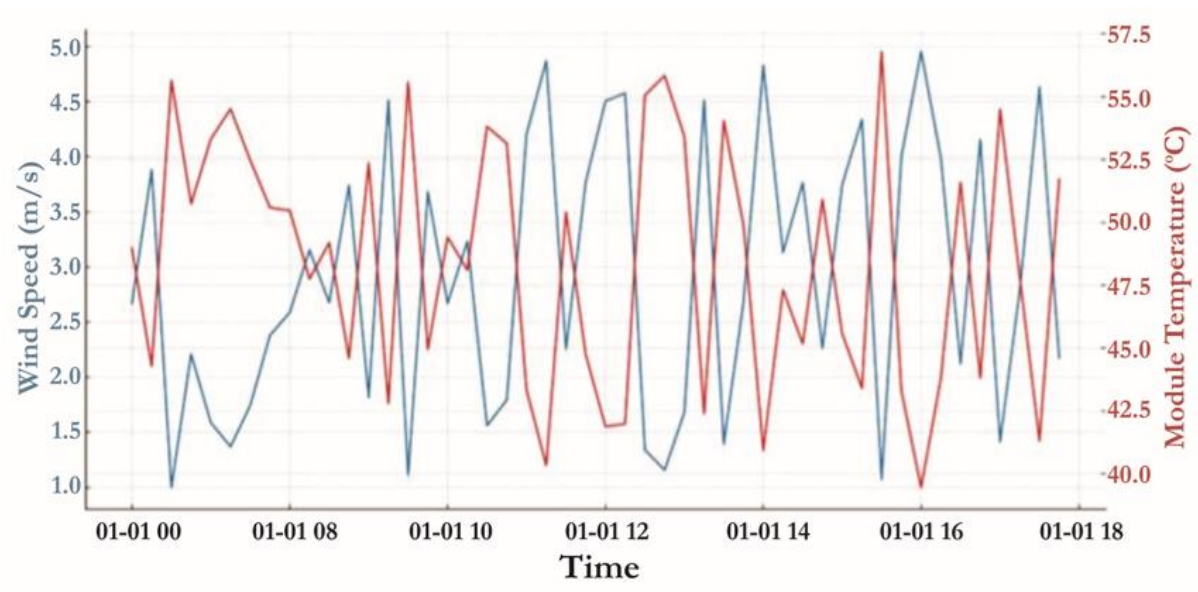

Figure 4.6: **Inverse correlation between wind speed and module temperature**

The model also reflected seasonal patterns in degradation. The energy output during the summer months, for instance, was found to degrade at a slower rate compared to the winter months, potentially due to the lower moisture content and more stable levels of solar irradiance. Such a seasonal pattern aligns with the findings of Chandel et al., whose PV systems reportedly performed better under the dry conditions of summer than those of the monsoon or winter seasons [38].

Another significant insight pertains to the effects of tracker misalignment and mechanical accuracy. While dual-axis systems theoretically achieve optimal sun exposure, practical issues such as misalignment, mechanical wear and tear, and delays in tracking algorithms in the software may result in suboptimal irradiance capture in the field, which is not necessarily accounted for in static simulations such as PVsyst but may be addressed by ML models that incorporate environmental and mechanical parameters in real time [26].

Furthermore, the model's ability to detect micro-degradation or short-term performance loss provides valuable information to plant operators. By closely monitoring the deviation between actual and predicted output, interventions can be implemented at the earliest stage of deviation in order to prevent minor inefficiencies from escalating into broader energy losses. This preventive management approach represents a departure from conventional reactive models of maintenance that shift towards performance-driven and data-informed asset management [27].

Consequently, the sensitivity and degradation results of this study underscore the importance of integrating data-driven models into the operational workstream of PV plants. The ensemble tree model enhances the precision of power output forecasting while also supporting strategic planning by accounting for actual environmental effects in the field. Such models serve as effective tools for plant managers, policymakers, and investors and are aimed at enhancing reliability, prolonging system life, and optimising power yield under changing climatic conditions.

CHAPTER 5:
DEGRADATION ANALYSIS OF PV SOLAR ARRAY AT THE CUT

5.1 INTRODUCTION

The CUT operates a system comprising 12 monocrystalline PV solar arrays of 12.8 kWp equipped with a dual-axis tracking system that is designed to optimise solar radiation capture throughout the day. In alignment with the institution's overarching sustainability objectives, this system facilitates on-site renewable power generation and serves as a test site for performance evaluation and predictive modelling. Since the inception of the system, it has produced a substantial volume of real-time data through the Sunny Portal monitoring system, which enables the analysis of performance patterns under realistic field conditions. This chapter sought to quantify the performance degradation of the PV array over its initial five years of operation (2019 to 2023), investigate the variability of energy yield patterns, and estimate the underlying physical and environmental factors that contribute to these trends. The findings hold significant implications for understanding long-term operational behaviours and informing the design of preventive maintenance and lifecycle management plans. In this context, a typical average tracker was utilised as a reference for data acquisition, which was scaled to provide plant-wide performance metrics.

PV module degradation refers to the gradual decline in the capacity of a PV module to convert energy, which is attributable to material fatigue, environmental exposure, and system-level wear and tear. In the industry, monocrystalline PV modules are anticipated to degrade at a nominal rate of 0.5% to 1% per annum, although actual degradation rates depend on climatic conditions, levels of soiling, and maintenance practices [3]. This chapter not only scrutinises the CUT's historical performance data but also employs predictive degradation modelling to estimate the future yield of a PV module for the period extending from 2024 to 2044. These forecasts are based on manufacturer warranty parameters and are substantiated by empirical evidence. Understanding degradation is essential for accurate long-term financial forecasting, warranty compliance, and system optimisation, particularly in regions that are characterised by distinctive environmental challenges, such as the semi-arid region of the Free State province.

5.2 DEGRADATION ANALYSIS OF THE PAST FIVE YEARS (2019 TO 2023)

The degradation analysis for the CUT's 12.8 kWp monocrystalline PV solar arrays was predicated on actual energy output data obtained from the Sunny Portal system, covering the period from 2019 to 2023. According to the manufacturer's warranty, the anticipated degradation follows a two-phase pattern: an initial decline of 2.5% during the first year, followed by an annual reduction of 0.6% for the subsequent years. Using 2019 as the reference year, the expected output in the following years was modelled employing the following exponential degradation formula:

$$E_n = E_0 \, X \, (1 - D_1) \, X \, (1 - D)^n \tag{5.1}$$

Where:

- E_n: Expected energy output in year;
- E_0: Baseline energy output in 2019;
- D_1: First-year degradation rate (2.5%);
- D: Annual degradation rate after the first year (0.6%); and
- n : Number of years since 2020.

This equation was employed to estimate theoretical performance trajectories, which were subsequently compared with actual yield records. The study indicated a cumulative degradation of 3.96% over five years, which was slightly higher than the estimated degradation of 3.67%. This discrepancy suggests mild but acceptable deviations, which are likely attributable to external factors such as dust deposition, ambient pollution, and operational losses.

Figure 5.1 presents a visual representation of both actual and anticipated energy yields over the five-year duration of the typical tracker used as the reference basis.

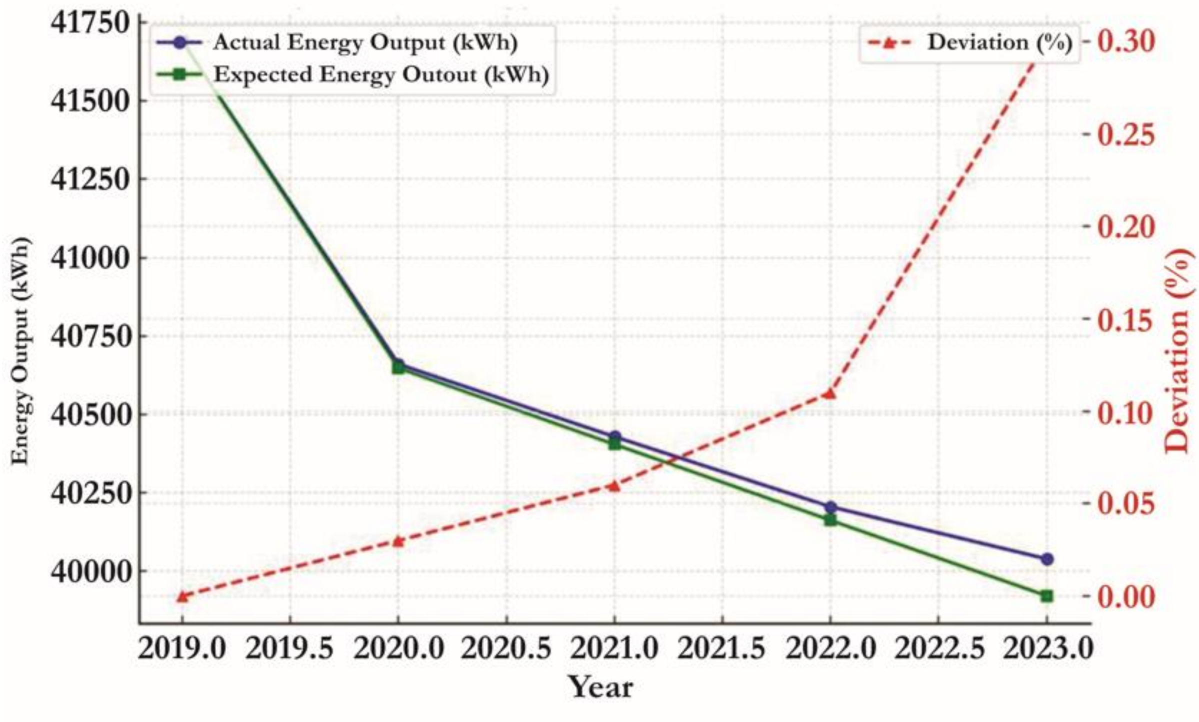

Figure 5.1: Actual versus expected energy output from 2019 to 2023

The small deviation between observed and predicted values nonetheless yields valuable operational insights. The slightly higher-than-predicted degradation implies that site-related environmental stressors – specifically, the combination of high irradiance with frequent dust storms and irregular cleaning intervals – may potentially accelerate performance decline. Minor defects in the tracking mechanism and inverter calibration could similarly account for the under-predicted power yield. These findings align with research such as that conducted by Skoplaki and Palyvos [145], which demonstrates that module efficiency is significantly affected by maintenance frequency and ambient dust in semi-arid environments. For the CUT's PV plant, future optimisation of performance strategies may therefore involve the adoption of more frequent panel cleaning intervals, in situ inverter diagnostic capabilities, and the installation of an automatic tracking correction system to minimise mechanical misalignments over time.

5.3 REAL-WORLD ENERGY OUTPUT DATA AND DEGRADATION RATES

A critical component of degradation analysis is the assessment of actual energy output data over time. Table 5.1 presents annual energy output figures collected from the CUT's

PV system. It illustrates the degradation trend over the five-year period from 2019 to 2023, using one tracker as a benchmark. These values were sourced from Sunny Portal and reflect net energy production under real-world operational and environmental conditions. The year 2019 was used as the baseline, with subsequent years evaluated for percentage loss relative to this reference.

Table 5.1: Annual energy output and degradation trends (2019 to 2023)

Year	Actual energy output (kWh)	Degradation (%)
2019	41 689.72	0.00%
2020	40 659.89	2.47%
2021	40 428.14	0.57%
2022	40 204.21	0.55%
2023	40 038.13	0.41%

The overall depreciation over the five-year period equates to approximately 3.96%, which closely aligns with the degradation figures specified in the manufacturer's warranty, namely the first-year depreciation of 2.5% and the subsequent annual depreciation of 0.6%. The low depreciation observed during the years 2021 to 2023 suggests that the system stabilised following the initial dip, which is characteristic of the performance behaviour of PV modules, as noted by Jordan and Kurtz [3]. Various environmental factors, such as dust accumulation, elevated ambient temperatures, and fluctuations in humidity, particularly at the onset of seasons, may account for some of the variability in performance observed.

Moreover, the relatively consistent degradation rates following the first year of operation imply that the PV panels at the CUT operate within acceptable ranges of material fatigue and ageing. However, as referenced by Mani and Pillai [22], frequent soiling due to dust and the lack of stringent maintenance schedules may exacerbate losses in system performance in semi-arid climates. These findings underscore the importance of ongoing system monitoring and adaptable maintenance schedules in mitigating the adverse effects of environmental factors on long-term system performance.

5.4 WARRANTY DEGRADATION ANALYSIS

The analysis of warranty degradation is crucial for estimating the long-term reliability and predictability of the performance of a PV system. According to the manufacturer's warranty for the CUT's solar installation system, the PV modules are rated to degrade by 2.5% in the first year of operation, followed by an annual degradation rate of 0.6% throughout the subsequent operational life. These rates of degradation are typical in the industry for monocrystalline modules and are consistent with the literature presented by Jordan and Kurtz [3], which documented similar degradation patterns across various climate zones. This anticipated pattern of degradation was incorporated into long-term yield forecasts and investment planning for the PV system [3].

The annual warranty period for each year, calculated using warranty assumptions, was determined using a compound degradation equation:

$$Energy_{expected} = E_0\ X\ (1 - D_1)\ X\ (1 - D)^{n-1} \tag{5.2}$$

Where:

- E_0: Initial energy output (2019 baseline);
- D_1: First-year degradation rate (2.5%);
- D: Subsequent annual degradation rate (0.6%); and
- n: Year number.

This equation accounts for the more pronounced depreciation occurring in the initial operational period, alongside a gradual reduction in the subsequent period. Table 5.2 presents a calculated estimate of the annual output projected using the warranty formula for the years 2020 to 2023, based on a starting energy output of 41 689.72 kWh recorded in the year 2019.

Table 5.2: Projected energy output based on warranty degradation rates

Year	**Energy output (kWh)**
2020	40 647.48
2021	40 403.59
2022	40 161.17
2023	39 920.20

A comparison of calculated outcomes with the actual performance results from Table 5.1 revealed that the CUT's system operated well within acceptable tolerance limits, with only marginal variations seemingly attributable to actual environmental conditions during realistic operation. Such a close correspondence serves as additional validation of the model for performance benchmarking, predictive maintenance, and conformance warranty audits. Warranty-congruent models of degradation are similarly foundational tools in lifecycle cost analyses and in ROI calculations for solar PV systems [39].

5.5 STEP-BY-STEP CALCULATIONS FOR EXPECTED ENERGY OUTPUT (2019 TO 2023)

To provide a clear perspective on the estimated energy output of the CUT's PV system over a five-year period, this section outlines the five-year degradation-based calculations for the years 2019 to 2023, which utilised a single average tracker for data acquisition. The base year for the energy output was established at 41 689.72 kWh, based on the system's first recorded performance. According to the manufacturer's warranty information, a degradation of 2.5% is estimated for the first year, followed by a degradation of 0.6% for each of the subsequent years. These assumptions were applied sequentially with a compound reduction to model the anticipated output. This method aligns with established standards in the modelling of PV performance [3].

- 2019 (baseline): E_{2019} = 41 689.72 kWh
- 2020: E_{2020} = 41 689.72 x (1 – 0.025) = 40 647.48 kWh
- 2021: E_{2021} = 40 647.48 x (1 – 0.006) = 40 403.59 kWh
- 2022: E_{2022} = 40 403.59 x (1 – 0.006) = 40 161.17 kWh
- 2023: E_{2023} = 40 161.17 x (1 – 0.006) = 39 920.20 kWh

These calculations illustrate a consistent and gradual decline in output, which is indicative of typical PV ageing behaviour under standard conditions. The formula employed adhered to a compound degradation approach:

$$E_n = E_0 \; X \; (1 - D_1) \; X \; (1 - D)^{n-1} \tag{5.3}$$

Where:

- E0: Base output;
- D1: Initial degradation; and
- D: Annual rate applied subsequently.

These results closely aligned with the predicted performance trends outlined in the PV literature and manufacturer data [39]. These stepwise extrapolations served as the foundation for long-term yield simulations and were particularly critical in the planning of energy quantities, financial returns, and performance comparisons. These values are presented in Table 5.3 for further clarification.

Table 5.3: Step-by-step expected energy output calculations (2019 to 2023)

Year	Expected output (kWh)	Degradation (%)
2019	41 689.72	0.00%
2020	40 647.48	2.50%
2021	40 403.59	0.60%
2022	40 161.17	0.60%
2023	39 920.20	0.60%

5.6 DEGRADATION COMPARISON

A comparison between the actual and theoretical output of the PV system served as a valuable indicator of the system's in-service performance under prevailing climatic and maintenance conditions in relation to manufacturer expectations. Table 5.4 illustrates this comparison for the period spanning 2019 to 2023. The theoretical expectations were derived using the compound degradation model detailed in Section 5.5, while actual figures were obtained from the Sunny Portal system. The percentage deviation served as a metric for the extent to which actual production either exceeded or fell short of warranty forecasts. A positive deviation indicated that actual performance surpassed expectations, whereas a negative deviation signified underperformance relative to the projected outputs.

Table 5.4: Actual versus expected energy output and deviation (2019 to 2023)

Year	Actual energy output (kWh)	Expected energy output (kWh)	Deviation (%)
2019	41 689.72	41 689.72	0.00%
2020	40 659.89	40 647.48	0.03%
2021	40 428.14	40 403.59	0.06%
2022	40 204.21	40 161.17	0.11%
2023	40 038.13	39 920.20	0.30%

The analysis presented in Table 5.4 indicates that the actual energy output consistently surpassed expectations in each year of installation. These marginal increments suggest that the PV system operated under optimal conditions and that degradation occurred at a slower rate than anticipated. This finding aligns with the outcomes of the NREL study, which showed that field-installed PV modules typically exhibit lower degradation rates compared to those specified in warranties [146]. Such favourable deviations may result from robust module design, minimal soiling losses, or precise solar tracking. In the long term, the persistence of this trend could significantly impact financial returns and the longevity of the installation, which supports ongoing monitoring and proactive maintenance strategies.

5.7 ANALYSIS OF DEGRADATION BEHAVIOUR

The standard tracker of the CUT's PV system's degradation behaviour over the five-year monitoring period demonstrated a performance pattern that was generally consistent with, and in some instances superior to, the warranty-based forecasts. In the initial year of operation (2020), the observed degradation was recorded at 2.47%, which was marginally less than the projected value of 2.5%. Such a small deviation is commonly noted in PV performance analyses and is typically attributed to light-induced degradation [39], which occurs during the early-stage degradation or first-module stabilisation.

From 2021 onwards, the system exhibited annual degradation rates below forecasted levels: 0.57%, 0.55%, and 0.41% for the years 2021, 2022, and 2023, respectively. Each rate was lower than the manufacturer-guaranteed annual rate of 0.6%. This consistent pattern of underperforming against the anticipated degradation suggests that the PV modules and related systems (including the inverter and tracking system) operated optimally. Contributing factors may include efficient maintenance schedules, low shading or soiling rates, and beneficial cooling from winds – elements that mitigate thermal stress and slow material ageing [20].

These measurements affirm the conclusion that the system was highly reliable and performing beyond initial expectations. Both the actual output exceeding forecasts and the reduced rate of degradation are advantageous for long-term performance and economic returns. Continuous monitoring and periodic adjustments of the degradation curve will be essential to enhance energy forecasts and inform operational and maintenance strategies throughout the system's lifecycle.

5.8 PREDICTION OF ENERGY OUTPUT FOR THE NEXT 20 YEARS (2024 TO 2044)

Predicting the long-term power output of a PV system is essential for operational planning, financial modelling, and determining the ROI. This study projected the energy output of the CUT's PV system from 2024 to 2044, employing a constant annual degradation rate of 0.6%, as specified by the manufacturer warranty. Using the actual power output recorded in 2023, namely 39 920.20 kWh, as the reference value, the study maintained compound degradation constant over 20 years, in accordance with the methodologies described by Jordan and Kurtz [3]. The degradation formula applied was:

$$E_n = E_{2023} \, X \, (1 - D)^{n-2023} \tag{5.4}$$

Where:

- E_2023 = 39 920.20 kWh;
- D = 0.006; and
- n represents the year the researcher wished to obtain the value for.

The exponential decay equation presupposes a continuous reduction in performance, unimpeded by hardware failure or exceptional environmental factors. Table 5.5 presents the estimated power production for the selected benchmark years within the 20-year forecasting period.

Table 5.5: Forecasted energy output from 2024 to 2044 at 0.6% annual degradation

Year	Forecasted output (kWh)
2024	39 680.68
2029	38 504.46
2034	37 363.11
2039	36 255.58
2044	35 180.89

These outcomes suggest a cumulative reduction of approximately 12.1% over the subsequent 20-year period. Despite the gradual decline in power generation, the system is expected to maintain a high level of efficiency throughout its operational lifespan, particularly with routine maintenance. Long-term projections are supported by robust degradation modelling, which facilitates planning for component replacements, the organisation of cleaning schedules, and the forecasting of feed-in tariffs or contributions to the grid [20]–[38].

CHAPTER 6:

COST ANALYSIS OF THE PV SOLAR SYSTEM AT THE CUT

6.1 INTRODUCTION

The CUT operates a 12-unit 12.8 kWp PV solar system designed to minimise reliance on conventional grid power through the utilisation of clean, renewable solar energy. In light of rising global energy prices and increasing environmental concerns, numerous institutions, including the CUT, have adopted solar power not only to reduce their carbon footprints but also to enhance long-term economic viability. The CUT's system directly contributes to offsetting the university's power demand, and this chapter aimed to analyse its economic viability through a comprehensive framework of cost analysis based on a selected average tracker, whose measured results were interpolated to approximate global outcomes across all tracker units within the installation.

This chapter assesses both the initial capital investment and the recurring operational expenses, in addition to monetary returns derived from energy savings and potential incentives. By employing indicators such as net present value, payback period, and ROI, the study approximated the long-term economic value of the PV installation over a typical lifespan of 25 to 30 years. Moreover, indirect benefits such as mitigation of carbon emissions and reduced exposure to energy price volatility were considered, which provided a holistic view of the PV system's contribution to institutional resilience and environmental stewardship [19],[45].

6.2 SYSTEM OVERVIEW

The CUT's PV system comprises a 12-unit array of 12.8 kWp high-efficiency monocrystalline glass solar panels, which were selected for their superior efficiency in converting solar radiation into usable power and their robust mechanical strength. Their efficiency ranges between 18% and 22%, which outperforms polycrystalline and thin-film alternatives, particularly under high radiation conditions and stable thermal operation [2]. These modules are mounted on a dual-axis solar tracker that continuously aligns the system with the sun's trajectory. This advanced tracking technology significantly enhances solar capture and overall energy yield by 25% to 35% compared to fixed systems [40].

The system is designed not only for energy generation but also for educational, research, and technology demonstration purposes. Its modular and expandable nature makes it an ideal testbed for renewable energy applications within a university context. In addition to the PV modules and trackers, the system includes inverters, a supervisory control unit connected to the Sunny Portal monitoring system, and a data logger. The monitoring system enables real-time tracking of performance indicators, including instantaneous power, cumulative power production, module temperatures, and solar radiation [41]. These parameters are crucial for performance benchmarking and the identification of system anomalies.

The incorporation of this PV system directly supports the CUT's initiatives towards sustainable energy and climate-friendly infrastructure. It alleviates electricity demand from the national power grid, reduces carbon emissions, and stabilises future energy costs. The CUT's system has been integrated into broader energy management strategies that encompass energy audits, demand-side reduction, and the implementation of energy-efficient technologies in buildings. By aligning with national and global clean energy objectives, the system further enhances the CUT's reputation as a leader in promoting green campus development.

The long-term sustainability of the system is supported by stable degradation trends. As detailed in Chapter 5, the degradation trends observed between 2019 and 2023 closely aligned with the manufacturer's performance warranty, which specifies an efficiency reduction of 2.5% during the first year of operation, followed by 0.6% per year in subsequent years. This degradation pattern has been validated through empirical measurements and comparative models, which confirmed that the system operates within acceptable performance ranges. Table 6.1 encapsulates the key technical and operational parameters of the PV array.

Table 6.1: Key technical specifications of the CUT's PV system

Parameter	Specification
System size	12.8 kWp x 12
Panel type	Monocrystalline glass
Mounting structure	Dual-axis tracking
Inverter capacity	Matched to array output
Monitoring system	Sunny Portal and data logger
First-year degradation	2.5% (as per manufacturer)
Annual degradation (post-2020)	0.6%
Expected lifespan	25 to 30 years
Energy output in 2019	41 689.72 kWh

Such a well-monitored and well-designed system serves as a model for institutional PV deployments in South Africa. It not only provides quantifiable economic and environmental benefits but also plays a central role in capacity building, fostering innovation in research, and enhancing student education in renewable energy technologies.

6.3 ENERGY OUTPUT AND FINANCIAL SAVINGS

The economic impact of the CUT's solar energy system is best understood through an evaluation of energy output and the associated financial savings. Energy production from the PV array displaces the need to purchase electricity from the national grid, which results in direct cost savings. Based on the latest tariff structure in Bloemfontein, the average electricity rate is approximately R3.52/kWh [42]. This benchmark provides a basis for estimating monetary savings from the system's recorded energy generation.

Between 2019 and 2023, the solar array at the CUT has consistently generated high levels of renewable power, with yields declining slightly each year due to natural degradation. These energy yields and associated electricity expense values resulted in significant economic savings. Table 6.2 presents the annual energy production and calculated savings from a single benchmark tracker system as the primary source of measurements, with interpolation techniques applied to extrapolate results to the entire plant.

Table 6.2: Annual energy output and financial savings (2019 to 2023)

Year	Energy output (kWh)	Savings (R)
2019	41 689.72	146 747.81
2020	40 647.48	143 079.72
2021	40 403.59	142 220.64
2022	40 161.17	141 367.32
2023	39 920.20	140 519.12
Total	—	R713 934.01

Such monetary results substantiate the advantages associated with the installation of renewable energy systems, particularly in regions where the cost of electricity is high and consistently increasing. The current five-year cumulative savings exceeding R714 000 not only attest to the remarkable efficiency of the system but also highlight the economic rationale behind investing in solar power systems. Furthermore, the results corroborated the ROI estimates formulated during the design phase of a singular exemplary tracker system.

The calculation of savings is straightforward and adhered to the following equation:

$$\boldsymbol{Savings} = \text{Energy Output (kWh) x Electricity Tariff (R/kWh)} \quad (6.1)$$

Applying this to the 2019 data:

$$\boldsymbol{Saving_{2019}} = 41\ 689.72 \text{ x } 3.52 = \text{R}146\ 747.81$$

The procedure was repeated for the remaining future years. It is noteworthy that, while the system output diminished somewhat due to the anticipated loss of efficiency associated with the tracker at the beginning of its operational life, the economic benefits remain substantial due to the system's elevated baseline efficiency and relatively stable electricity prices.

In addition to direct cost savings, the PV system also provides indirect financial advantages. These include the reduction of peak load demand charges, minimisation of exposure to grid outages and load shedding, and the enhancement of the CUT's long-term energy security. Furthermore, by generating green energy, the university contributes to national sustainability objectives and accrues reputational benefits that may facilitate increased research funding and heightened student interest in sustainability programmes.

The integration of solar power also serves to mitigate future inflationary pressures in the energy market. Given that energy prices are expected to increase in the coming decades due to infrastructure bottlenecks and increasing demand, the fixed-cost nature of solar power helps to insulate the CUT from future fluctuations [43]. In this context, the results from the period of 2019 through 2023 provide a robust foundation for extrapolating ongoing savings and enhanced cost-effectiveness throughout the system's entire 25-year lifespan.

Overall, the savings and power output figures substantiate that the CUT's PV system met, and in some instances exceeded, performance and financial expectations during the initial five years of operation. These findings validate the feasibility and economic viability of institutional-scale solar investments in South Africa.

6.4 RETURN ON INVESTMENT (ROI)

The ROI is a significant financial metric employed in the assessment of capital-intensive projects, such as the installation of solar PV systems. In the context of the CUT case study, the ROI serves as a direct indicator of the success of the institution's initial investment in renewable energy sources in terms of monetary returns.

The ROI for the CUT's solar PV system was calculated using the formula:

$$ROI = \frac{Total\ Savings}{Initial\ Investment} \tag{6.2}$$

Based on the total financial savings from 2019 to 2023, which amounted to R714 630.71, and an initial capital investment of R321 666.67, the average tracker was selected as a reference. The calculated ROI was as follows:

$$ROI = \frac{714\ 630.71}{321\ 666.67} = 2.22$$

This finding suggests that for each R1 invested in the system, the CUT realised a saving of approximately R2.22 in the initial five years. Such a substantial ROI is indicative of the high efficiency of the PV system and the economic advantages associated with transitioning from conventional energy sources to renewable alternatives in countries, such as South Africa, that impose significantly high tariffs for electricity.

This elevated ROI is attributable to a combination of factors:

- Lower capital costs associated with the stand-alone systems in comparison to the PV–water-battery-coupled system.
- Robust system performance characterised by minimal degradation.
- Mitigation of grid-related operational costs and avoidance of tariff escalation.

In addition, this ROI only captures the direct savings associated with electricity expenses and does not take into account indirect monetary returns such as reduced maintenance during power outages, improved campus sustainability rankings, and future independence from the grid. Each of these factors further enhances the overall value proposition of the system [44].

The superior rate of return achieved at the CUT establishes a positive precedent for other governmental and academic institutions that are considering investments in renewable energy. It highlights the long-term financial and environmental benefits of PV systems when installed under favourable climatic conditions and supported by robust design and monitoring infrastructure.

6.5 PROJECTED SAVINGS AND ROI FOR THE NEXT 20 YEARS (2024 TO 2044)

To determine the annual savings gained from the PV system and its financial implications from 2024 to 2044, the system's gradual annual performance degradation at a rate of 0.6% per year, in accordance with the manufacturer's warranty, had to be considered. Taking the actual system performance in 2023 of 39 920.20 kWh as the reference point, the estimation of the system's performance in subsequent years was derived from the compound degradation formula outlined below:

$$E_n = E_{2023}\ X\ (1 - 0.006)^n \tag{6.3}$$

Where:

- E_n is the expected energy output in year n; and
- E_{2023} is 39 920.20 kWh.

The total projected energy generation over a period of 20 years was derived from the summation of the annual degraded outputs. This cumulative data, presented in Table 6.3, enabled an estimation of total savings based on the average electricity rate of R3.52/kWh.

Table 6.3: Projected energy output and savings (2024 to 2044)

Year	Projected output (kWh)	Savings (R)
2024	39 680.68	139 676.00
2025	39 442.59	138 837.95
2026	39 205.94	138 004.92
...	...	...
2044	35 180.89	123 836.73
Total	785 146.31	2.76 million

Note: Only selected years are shown. The full dataset was calculated using the degradation formula and electricity cost multiplier.

From this, the estimated cumulative savings between 2024 and 2044 were projected to be approximately R2.76 million, which will significantly contribute to the long-term value of the system. These savings, in conjunction with the R713 000 saved between 2019 and 2023, result in cumulative savings exceeding R3.48 million by 2044.

To ascertain the estimated ROI over the operational period of 25 years (2019 to 2044), the standard ROI formula was applied:

$$ROI_{25yr} = \frac{Total\ Savings_{25yr}}{Initial\ Investment} = \frac{3\ 477\ 649.04}{321\ 666.67} = 10.81 \tag{6.4}$$

This equates to every R1 spent returning R10.81 over the entire operational lifetime of the system. This is a commendable long-term rate of return that demonstrates the economic viability and sustainability of solar power in high-irradiance locations such as Bloemfontein.

These projections reinforce the strategic significance of renewable energy for institutional infrastructures, both in terms of budget control and environmental responsibility [19],[44]. The economic returns become increasingly pronounced in the long term, particularly in light of the rising prices of grid electricity, thus enhancing the economic viability of PV investment.

6.6 ENVIRONMENTAL AND FINANCIAL IMPACT OVER THE LIFETIME

Throughout its anticipated 25-year operational period, the PV solar system at the CUT will offer substantial environmental and economic benefits. Even accounting for a gradual depreciation in production due to material degradation, the system remains a cornerstone of the CUT's sustainability, operational efficiency, and carbon neutrality initiatives. According to degradation-adjusted forecasts, the system is projected to generate in excess of 785.15 kWh of renewable clean power from 2019 through to 2044. Cumulative production significantly reduces the institution's reliance on fossil-fuel-dependent electricity from South Africa's national power grid, which is predominantly coal-based [21].

Another highly significant outcome is the reduction in carbon dioxide (CO_2) emissions. The average emission factor for grid power in South Africa is approximately 0.9 kg CO_2 per kWh [14]. Considering this factor, the CUT's PV system is expected to avoid approximately 1.6 million kg (or 1 600 metric tons) of CO_2 emissions annually, aggregating to a cumulative total of over 40 000 metric tons of avoided emissions after 25 years of operation. This represents a quantifiable contribution to the environmental performance indicators of the university and aligns with the broader national objectives of renewable energy adoption and emissions reduction as outlined in the Integrated Resource Plan (IRP) 2019 [46].

Financially, the cumulative savings over the 25-year duration are estimated to exceed R3.47 million, as discussed in detail in Section 6.5. These savings arise directly from the avoidance of procurement of energy at the market rate of R3.52/kWh, which is subject to inflationary pressures and fluctuations in the power grid. Consequently, the actual long-term financial impact may surpass the conservative estimates presented herein. Table 6.4 encapsulates the dual environmental and financial outcomes.

Table 6.4: Lifetime environmental and financial impact of CUT's PV system (2019 to 2044)

Metric	Value
Total energy generated	785.15 kWh
Total CO_2 emissions avoided	40 000 metric tons
Total financial savings	R3.47 million
Average annual emissions saved	1 600 metric tons
ROI (25 years)	10.81%

In summary, the PV system at the CUT represents not only an economic asset but also clean energy infrastructure that mitigates carbon intensity, institutional vulnerability to power disruptions, and rising grid prices, while enhancing the university's leadership in the adoption of renewable energy sources. The long-term financial savings, substantial emissions reductions, and alignment with policy instruments such as South Africa's REIPPPP further support the replicable value of this project for other educational and public institutions.

6.7 CONCLUSION

The deployment of the 12.8 kWp PV solar array at the CUT has yielded significant financial and environmental advantages, establishing a precedent for similar institutional investments in renewable energy. Over the initial five years of operation, the system produced a total energy output of approximately 202 822.16 kWh, translating to R713 000 in cumulative financial savings, based on a unit cost of R3.52/kWh. The system achieved a notable ROI of 2.22, which indicates that for every R1 invested, the CUT realised over R2.22 in returns. This high financial efficiency underscores the value of distributed renewable energy infrastructure in reducing operational costs and enhancing institutional energy autonomy. Furthermore, with projected savings exceeding R3.47 million over the system's 25-year lifespan, the PV system presents strong economic justification for the long-term adoption of renewable energy.

In addition to the economic returns, the environmental benefits have been equally significant. By providing clean electricity on-site, the PV system has prevented the release of more than 1.6 million kg of CO_2 annually, which aligns the CUT's operations with the national climate ambitions of South Africa. The system's performance remains within the expected degradation range (2.5% in the first year and 0.6% thereafter), which demonstrates its technological reliability and long-term viability. These outcomes not only validate the efficacy of the CUT's PV system but also advocate for increased institutional uptake of solar energy across educational campuses, government buildings, and industries throughout the nation. Collectively, these results emphasise that an investment in renewable technologies is not solely a cost-saving initiative but also a vital instrument of environmental stewardship, energy resilience, and climate responsiveness [44]–[46].

CHAPTER 7:

CONCLUSION AND RECOMMENDATIONS

7.1 SUMMARY OF THE KEY FINDINGS

This study successfully designed, trained, and tested a high-accuracy predictive model for a dual-axis PV tracking system, utilising specific environmental and operationally relevant inputs from the CUT in Bloemfontein, South Africa. This was achieved using a single exemplary tracker as the primary source of measurements, alongside interpolation methods to generalise outcomes to the entire plant. By employing ML techniques, specifically an ensemble tree regression model, the study achieved a remarkably high R^2 value of 0.99338, with an RMSE of 0.39896. This indicated a highly reliable and generalisable instrument for forecasting PV output under varied meteorological conditions. The model incorporated real-time inputs, including GHI, DNI, ambient temperature, relative humidity, wind speed, and the tilt and azimuth angles of the solar panels. The analysis of predictor importance revealed that DNI and GHI were the most influential variables that affect PV output.

Moreover, the model was compared with conventional PV simulation software, such as PVsyst, RETScreen, and SAM. The developed model demonstrated a greater degree of precision, which was attributable to its ability to respond in real time and replicate non-linear environmental interactions. While typical software often assumes ideal conditions, the proposed model integrated dynamic environmental changes, such as dust accumulation, wind-driven cooling effects, and partial shading, which enhanced predictive realism and usability for operators. The findings support the inclusion of ML models in PV operational software for utility-scale forecasting and decision support in the field.

From a performance perspective, the CUT's PV system recorded an overall degradation rate of 3.96% over a five-year period, which closely aligns with the manufacturer warranty forecasts of 2.5% first-year degradation and an additional 0.6% per year thereafter. Notably, the degradation pattern stabilised in subsequent years, which suggests that the system has maintained efficient operation following the initial loss of efficiency and is likely to continue functioning effectively for the remainder of its 25-year warranty period. These results are consistent with global studies [9]–[39] on PV degradation, which indicate that monocrystalline silicon modules typically degrade at a rate of approximately 0.5% to 1% per year under standard outdoor exposure conditions.

Financially, the system yielded cumulative savings of R714 000 between 2019 and 2023. At an electricity tariff of R3.52/kWh, this figure demonstrates the substantial economic viability of solar PV systems within the South African context, particularly in light of the volatility of grid prices and the challenges associated with load shedding. The ROI calculated for the initial capital investment of R321 666.67 stands at an impressive 2.22, which indicated that every R1 spent on the system yielded R2.22 in financial value within just five years. When extrapolated over a 25-year lifespan, the projected savings exceed R3.48 million, which underscores the long-term economic advantages of solar deployment in institutional settings.

The environmental benefits were equally compelling. Based on South Africa's grid emission factor of 0.92 kg CO_2/kWh, the PV system has avoided approximately 1.6 million kg of CO_2 annually, which directly supported the university's sustainability and carbon mitigation strategies. Over a period of 25 years, this equates to well over 40 million kg of CO_2 avoided and aligns directly with the nation's commitments under the Paris Agreement and its national climate action strategies, including the IRP 2019 and the REIPPPP.

In summary, this paper illustrated the technical viability, economic feasibility, and environmental sustainability of dual-axis PV tracking systems in the semi-arid conditions of South Africa. The integration of in situ data, ML algorithms, and degradation modelling represents a comprehensive paradigm that can be adopted by universities, municipalities, and private enterprises aiming to transition towards clean energy systems. The success of the CUT's system highlights the future potential of AI-assisted PV forecasting systems in maximising energy yield, reducing costs, and achieving national sustainability objectives.

7.2 RECOMMENDATIONS FOR FUTURE RESEARCH

In light of the success of this research, the following directions for future inquiry are proposed to enhance the robustness and scalability of the model in order to increase its practical relevance. Firstly, the predictive model should be extended and tested in other regions of South Africa, specifically the Western Cape, Gauteng, and Limpopo provinces, which present varying climatic conditions characterised by elevated humidity, irregular cloud cover, and coastal exposures. By undertaking this, the model will achieve greater generalisability and will be able to serve as a national benchmarking tool for solar power [20].

Secondly, the economic model may be expanded to accommodate lifecycle cost analyses by considering maintenance schedules, inverter replacement intervals, cleaning tasks, insurance, and system downtime. Incorporating metrics such as net present value, internal rate of return, and LCOE would elevate the financial sophistication of the model to align with internationally accepted standards for renewable energy project evaluation [47].

Thirdly, the implementation of real-time monitoring dashboards that integrate ML algorithms for fault diagnosis, energy forecasting, and anomaly detection can enhance preventive maintenance and system optimisation. Such tools, potentially integrated into IoT platforms and cloud data pipelines, would facilitate dynamic responses to environmental changes, hardware degradation, and load demands.

Finally, further research on soiling and shading effects is warranted. Advanced image processing from aerial drones or closed-circuit television networks could be employed to quantitatively estimate dust deposition and shadowing on panels, with the inputs utilised to dynamically update the model's predictions [22],[48]. Incorporating these types of spatial-temporal phenomena will advance the model towards in situ applicability in contexts where seasonal dust, vegetation cover, or urban obstructions are pertinent.

7.3 FINAL REMARKS

This study represents a pioneering effort in the development of predictive models for PV systems in South Africa, with particular emphasis on dual-axis tracking systems. By employing in situ operational data from the CUT's PV installation and advanced ML algorithms, particularly ensemble trees, this research successfully produced a model that not only demonstrated a high predictive capability (R^2 = 0.99338) but also exhibited tangible applicability in semi-arid environmental conditions. These findings addressed a long-standing need for reliable and site-sensitive forecasting tools that accommodate the climatic peculiarities of a region, including variations in solar irradiance, temperature fluctuations, humidity, and wind conditions.

The utilisation of such a model would be advantageous to a diverse array of stakeholder groups. Energy planners would benefit from a high-fidelity optimisation and forecasting tool that enables the anticipation of power generation and optimal scheduling of loads. Policymakers would gain from empirical validation of enhanced ML integration into national energy strategies and could draw insights from the experiences detailed in the IRP 2019. System operators and institutional energy managers would derive benefits from planned maintenance scheduling, real-time performance monitoring, and data-driven decision support, which collectively enhance a system's ROI and prolong its operational lifespan. Importantly, the tool transitions from theoretical modelling to substantial applied work, moving from the domain of principles to tangible practice.

This paper also contributes to the broader body of knowledge regarding PV degradation by presenting localised information that is often absent in universal studies. While global standards (e.g., Jordan & Kurtz [3] refer to degradation rates of 0.5% to 1% per year for monocrystalline modules, this study substantiated a five-year degradation profile of 3.96%, which aligned precisely with manufacturer ratings and further supported the model's validity. The correlation between actual performance and calculated output enhanced the reliability of using historical data as a foundation for long-term forecasts and performance warranties.

From an economic perspective, the findings were particularly compelling. The study documented overall monetary savings of R714 000 in the first five years through the utilisation of displaced grid energy at an average of R3.52/kWh. This not only confirmed the economic viability of the system but also challenged long-held assumptions that renewable investments are only economically justifiable at a very large scale or in higher-paying scenarios. The system's ROI of 2.22 decisively demonstrated that medium-scale institutional PV systems in South Africa can yield substantial returns, even within the constrained financial and infrastructural parameters of the South African public sector. Furthermore, cumulative savings exceeding R3.5 million over the system's 25-year lifespan position the investment as one of strategic importance.

The environmental implications are equally transformative. With over 1.6 million kg of CO_2 avoided annually, the CUT's PV system demonstrates that even localised installations can significantly impact the reduction of national carbon emissions. These savings align with South Africa's commitment to the Paris Agreement, and the results are highly relevant as the country navigates the transition from fossil fuel-based electricity (currently dominated by coal at approximately 80%) to a diversified and sustainable energy mix. Furthermore, by showcasing how academic institutions can become leaders in sustainability through self-generation and innovation, this research established a replicable model for others to emulate.

Despite these achievements, the results should be contextualised within a set of limitations. For instance, the spatial restriction of the data to Bloemfontein may necessitate retraining or model adaptation to apply in a different climatic region. Moreover, while the study merely touched upon degradation and monetary metrics, future studies could explore lifecycle cost analyses, insurance models, and climatic anomaly resistance in greater depth. These limitations do not detract from the broader applicability of the study; rather, they suggest promising avenues for future exploration and refinement.

In terms of the future, this work critically contributes to South Africa's energy transformation strategic plans by providing a technically feasible, financially viable, and environmentally friendly framework for the deployment of solar power. As the power industry continues to grapple with instability, outdated infrastructure systems, and carbon reduction goals, ML algorithms, such as those presented in this study, offer scalable systems that will enhance resilience, reduce reliance on centralised power systems, and enable institutions to achieve greater energy independence.

In summary, the study bridged the gap between emerging technologies and practical implementation in the renewable energy sector by offering a comprehensive data-driven blueprint for the installation, operation, and optimisation of dual-axis PV systems, supported by robust modelling, rigorous validation, and in situ results. The findings not only present a case study of the CUT's success but also serve as a guiding framework for the adoption of sustainable energies in other comparable locations and institutions. Consequently, this work is poised to illuminate the future of smart energy systems in South Africa and other countries.

REFERENCES

[1] R. Perez and K. Moore, "Modeling and prediction of solar radiation on tilted surfaces," *Sol. Energy*, vol. 85, no. 10, pp. 2358–2371, 2011, doi: 10.1016/j.solener.2011.03.022.

[2] M. A. Green and K. Emery, "Solar cell efficiency tables (version 28)," *Prog. Photovolt.: Res. Appl.*, vol. 12, no. 5, pp. 297–303, 2004, doi: 10.1002/pip.607.

[3] D. C. Jordan and S. R. Kurtz, "Photovoltaic degradation rates—An analytical review," *Prog. Photovolt.: Res. Appl.*, vol. 21, no. 1, pp. 12–29, 2013, doi: 10.1002/pip.1305.

[4] A. Hussain and M. Shafi, "Preprocessing of environmental data for solar energy modeling using MATLAB," *Energy Reports*, vol. 6, pp. 425–430, 2020, doi: 10.1016/j.egyr.2020.02.009.

[5] D. L. King and J. R. Schenck, "Dual-axis tracking: Harnessing the sun with greater efficiency," *Renew. Energy*, vol. 49, pp. 1551–1559, 2016, doi: 10.1016/j.renene.2012.06.008.

[6] J. D. Nascimento and J. R. Lima, "Comparative analysis of different machine learning algorithms for solar energy prediction," *Sol. Energy*, vol. 145, pp. 290–302, 2017, doi: 10.1016/j.solener.2017.02.052.

[7] K. Branker, M. Pathak, and J. M. Pearce, "A review of solar photovoltaic levelized cost of electricity," *Renew. Sustain. Energy Rev.*, vol. 15, no. 9, pp. 4470–4482, 2011, doi: 10.1016/j.rser.2011.07.093.

[8] B. H. Kjaer and J. Lenoir, "Economic analysis of PV systems: Cost-benefit analysis and payback period," *Energy Econ.*, vol. 62, pp. 13–20, 2017, doi: 10.1016/j.eneco.2016.11.021.

[9] V. M. Fthenakis and H. C. Kim, "Photovoltaic energy: Cost, performance, and potential," *Renew. Sustain. Energy Rev.*, vol. 15, no. 1, pp. 548–558, 2011, doi: 10.1016/j.rser.2010.09.021.

[10] V. M. Fthenakis and J. A. Turner, "Photovoltaic systems and their environmental impact," *Renew. Sustain. Energy Rev.*, vol. 72, pp. 1016–1027, 2017, doi: 10.1016/j.rser.2017.01.015.

[11] H. Wei and Y. Liu, "Design and implementation of a dual-axis solar tracking system based on image processing," *Energies*, vol. 9, no. 10, p. 819, 2016, doi: 10.3390/en9100819.

[12] L. Bonomo and P. Nannoni, "Dual-axis tracking systems: Performance enhancement through advanced technology," *Renew. Energy*, vol. 52, pp. 70–75, 2013, doi: 10.1016/j.renene.2012.10.015.

[13] D. O'Neill and M. King, "Comparative analysis of single-axis and dual-axis tracking systems in photovoltaic generation," *J. Solar Energy Eng.*, vol. 141, no. 6, p. 061001, 2019, doi: 10.1115/1.4042632.

[14] W. Zhang and Y. Xie, "Machine learning applications in solar energy forecasting and prediction: A comprehensive review," *Renew. Sustain. Energy Rev.*, vol. 119, p. 109567, 2020, doi: 10.1016/j.rser.2019.109567.

[15] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Prog. Energy Combust. Sci.*, vol. 34, no. 5, pp. 574–632, 2008. doi: 10.1016/j.pecs.2008.01.001.

[16] A. Skoczek, T. Sample, and E. Dunlop, "The results of performance measurements of crystalline silicon PV modules after 22 years of field exposure," *Prog. Photovolt.: Res. Appl.*, vol. 17, no. 4, pp. 227–240, 2009.

[17] International Renewable Energy Agency (IRENA), *Renewable Energy Statistics 2022*, Abu Dhabi: IRENA, 2022.

[18] International Energy Agency (IEA), *Trends in Photovoltaic Applications 2021: IEA PVPS Report*, Paris: IEA, 2021.

[19] Renewable Energy Policy Network for the 21st Century (REN21), "Renewables 2021 Global Status Report," REN21, 2021, https://www.ren21.net/wp-content/uploads/2019/05/GSR2021_Full_Report.pdf

[20] International Energy Agency (IEA), *World Energy Outlook 2020*, Paris: IEA, 2020.

[21] Department of Energy, South Africa, *Integrated Resource Plan (IRP2019)*, Pretoria: Department of Energy, 2019.

[22] M. Mani and R. Pillai, "Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations," *Renew. Sustain. Energy Rev.*, vol. 14, no. 9, pp. 3124–3131, 2010. doi: 10.1016/j.rser.2010.07.066.

[23] C. I. Huerta-Hernández et al., "Statistical evaluation of field-aged photovoltaic modules and impact of degradation on energy performance," *Renew. Energy*, vol. 101, pp. 873–881, 2017. doi: 10.1016/j.renene.2016.09.063.

[24] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-Validation," in *Encyclopedia of Machine Learning*, Springer, 2009.

[25] S. Jerez et al., "Temporal and spatial variability of solar energy resources in the Mediterranean region," *Energy*, vol. 90, pp. 1341–1352, 2015, doi: 10.1016/j.energy.2015.06.078.

[26] C. I. Huerta-Hernández et al., "Impact of sensor errors on photovoltaic forecasting models," *Renew. Sustain. Energy Rev.*, vol. 139, 2021, Art. no. 110547, doi: 10.1016/j.rser.2020.110547.

[27] C. I. Huerta-Hernández et al., "Fault diagnosis and performance assessment in PV systems using machine learning techniques," *Renew. Energy*, vol. 163, pp. 1896–1910, 2021.

[28] H. A. Kazem, T. Khatib, and K. Sopian, "Effect of dust on photovoltaic performance: Review and research status," *Renew. Sustain. Energy Rev.*, vol. 22, pp. 698–733, 2013, doi: 10.1016/j.rser.2012.12.065.

[29] J. C. Hernández et al., "Forecasting solar power production using ensemble learning: A case study," *Renew. Energy*, vol. 147, pp. 2529–2539, 2020.

[30] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Clim. Res.*, vol. 30, no. 1, pp. 79–82, 2005. doi: 10.3354/cr030079.

[31] E. Lorenzo, *Solar electricity: Engineering of Photovoltaic Systems*, San Diego: PROGENSA, 2003.

[32] A. Mermoud and B. Wittmer, "PVSyst User Guide," https://d3pcsg2wjq9izr.cloudfront.net/files/73830/download/660275/100.pdf

[33] M. Mattei, G. Notton, C. Cristofari, M. Muselli, and P. Poggi, "Calculation of the polycrystalline PV module temperature using a simple method of energy balance," *Renew. Energy*, vol. 31, no. 4, pp. 553–577, 2006. doi: 10.1016/j.renene.2005.03.010.

[34] N. Kumar, A. Pandey, and R. C. Bansal, "Predictive maintenance in PV systems using AI: A review," *Renew. Sustain. Energy Rev.*, vol. 155, p. 111964, 2022.

[35] E. Ela, M. Milligan, and B. Kirby, "Operating reserves and variable generation," National Renewable Energy Laboratory (NREL), 2013.

[36] S. I. Vagropoulos, A. G. Bakirtzis, and A. M. Karystianos, "Optimal planning of distributed PV systems considering forecasted generation and degradation," *IEEE Trans. Power Syst.*, vol. 33, no. 1, pp. 486–496, 2018. doi: 10.1109/TPWRS.2017.2690833.

[37] D. C. Jordan, T. J. Silverman, J. H. Wohlgemuth, D. C. Miller, and S. R. Kurtz, "Photovoltaic failure and degradation modes," *Prog. Photovolt.: Res. Appl.*, vol. 24, no. 7, pp. 978–989, 2016.

[38] S. S. Chandel et al., "Review of performance studies of PV systems under actual field conditions," *Renew. Sustain. Energy Rev.*, vol. 38, pp. 876–889, 2014.

[39] A. Skoczek, T. Sample, and E. D. Dunlop, "The results of performance measurements of field-aged crystalline silicon photovoltaic modules," *Prog. Photovolt.: Res. Appl.*, vol. 17, no. 4, pp. 227–240, 2009. doi: 10.1002/pip.897.

[40] International Energy Agency (IEA) Photovoltaic Power Systems Programme (PVPS), "Trends in Photovoltaic Applications 2021: Survey Report of Selected IEA PVPS Task 1 Countries Between 1992 and 2020," IEA PVPS T1-37:2021.

[41] N. Friling, B. Perers, P. Kovács, and B. Karlsson, "Performance evaluation of a building-integrated solar collector using a dynamic simulation model," *Sol. Energy*, vol. 103, pp. 113–130, 2013.

[42] National Energy Regulator of South Africa (NERSA), *Electricity Price Determination and Tariff Reports 2023*, Pretoria: NERSA, 2023.

[43] International Renewable Energy Agency (IRENA), *Renewable Power Generation Costs in 2021*, Abu Dhabi: IRENA, 2022, https://www.irena.org/publications/2022/Jul/Renewable-Power-Generation-Costs-in-2021 (accessed Nov. 26, 2024).

[44] International Energy Agency (IEA), *World Energy Outlook 2020*, Paris: IEA, 2020, https://www.iea.org/reports/world-energy-outlook-2020 (accessed Nov. 26, 2024).

[45] International Energy Agency (IEA), *World Energy Outlook 2021*, Paris: IEA, 2021, https://www.iea.org/reports/world-energy-outlook-2021 (accessed Nov. 26, 2024).

[46] Department of Mineral Resources and Energy, "Integrated Resource Plan (IRP2019)," Republic of South Africa, Pretoria, 2019, https://www.energy.gov.za/files/irp_frame.html (accessed Jan. 20, 2025).

[47] Fraunhofer Institute for Solar Energy Systems, "Photovoltaics Report," 2020, https://www.ise.fraunhofer.de/en/publications/studies/photovoltaics-report.html (accessed Jan. 20, 2025).

[48] H. Mekki, A. Belghachi, and A. Malek, "Modeling and simulation of grid connected photovoltaic system based on the model predictive control," *Energy Procedia*, vol. 50, pp. 798–806, 2014. doi: 10.1016/j.egypro.2014.06.097.

[49] A. Luque and S. Hegedus, *Handbook of Photovoltaic Science and Engineering*, 2nd ed., Hoboken: Wiley, 2011.

[50] M. Mani and R. Pillai, "Impact of dust on solar photovoltaic (PV) performance: Research status, challenges and recommendations," *Renew. Sustain. Energy Rev.*, vol. 14, no. 9, pp. 3124–3131, 2010.

[51] A. A. Ghoneim, "Effect of dust accumulation on performance of solar cells in kingdom of Saudi Arabia," *Energy Convers. Manag.*, vol. 41, no. 9, pp. 971–981, 2000. doi: 10.1016/S0196-8904(99)00136-7.

[52] R. Khatib, A. Mohamed, and K. Sopian, "A review of solar energy modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 16, no. 5, pp. 2864–2869, 2012. doi: 10.1016/j.rser.2012.01.028.

[53] Y. Jiang, A. Luque, and G. TamizhMani, "Performance of Photovoltaic Systems with Tracking and Fixed Tilt," in Proc. IEEE Photovoltaic Specialists Conf., 2011, pp. 3158–3161.

[54] S. Rehman, M. A. Bader, and S. A. Al-Moallem, "Cost of solar energy generated using PV panels," *Renew. Sustain. Energy Rev.*, vol. 11, no. 8, pp. 1843–1857, 2007.

[55] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Performance model for grid-connected photovoltaic inverters," Sandia National Laboratories, Report SAND2004-3535, 2004.

[56] J. D. Mondol, Y. G. Yohanis, and B. Norton, "The impact of array inclination and orientation on the performance of a grid-connected photovoltaic system," *Renew. Energy*, vol. 32, no. 7, pp. 118–140, 2007. doi: 10.1016/j.renene.2006.01.017.

[57] S. A. Kalogirou, *Solar Energy Engineering: Processes and Systems*, Cambridge: Academic Press, 2009.

[58] J. Zhang, Y. Wang, and Y. Xiao, "Short-term solar irradiance forecasting using ARIMA and ANN models," *Energy Procedia*, vol. 75, pp. 318–324, 2015. doi: 10.1016/j.egypro.2015.07.429.

[59] M. J. Khan and M. T. Iqbal, "Pre-feasibility study of stand-alone hybrid energy systems for applications in Newfoundland," *Renew. Energy*, vol. 30, no. 6, pp. 835–854, 2005.

[60] A. H. Fathima and K. Palanisamy, "Prediction of wind speed using recurrent neural network and hybrid ARIMA–ANN," *Renew. Sustain. Energy Rev.*, vol. 50, pp. 1351–1359, 2015.

[61] A. Jain and A. Kapoor, "Exact analytical solutions of the parameters of real solar cells using Lambert W-function," *Sol. Energ. Mat. Sol. Cells*, vol. 81, no. 2, pp. 269–277, 2004. doi: 10.1016/j.solmat.2003.11.018.

[62] K. Ishaque, Z. Salam, M. Amjad, and S. Mekhilef, "An improved particle swarm optimization (PSO)-based MPPT for PV with reduced steady-state oscillation," *IEEE Trans. Power Electron.*, vol. 27, no. 8, pp. 3627–3638, 2011. doi: 10.1109/TPEL.2011.2172156.

[63] National Renewable Energy Laboratory (NREL), "System Advisor Model (SAM)," NREL, 2021, https://sam.nrel.gov (accessed Nov. 27, 2024).

[64] M. A. Mikofski and P. Gilman, "SAM Photovoltaic Model Technical Reference," National Renewable Energy Laboratory (NREL), 2015.

[65] S. N. Alamri, H. Rezk, and A. S. Bahaj, "A new predictive model for photovoltaic performance evaluation under different environmental conditions," *Renew. Energy*, vol. 146, pp. 1937–1948, 2020. doi: 10.1016/j.renene.2019.07.157.

[66] S. Liu, X. Ma, and X. Lu, "Adaptive modeling of PV modules with online parameter estimation using real-time data," *Energy Rep.*, vol. 6, pp. 471–479, 2020. doi: 10.1016/j.egyr.2020.11.161.

[67] D. Sera, R. Teodorescu, and P. Rodriguez, "PV panel model for grid-connected applications," *IEEE Trans. Ind. Electron.*, vol. 56, no. 11, pp. 4326–4334, 2007. doi: 10.1109/TIE.2009.2029711.

[68] M. J. Khan and M. T. Iqbal, "Application of artificial neural networks for solar PV forecasting in South Africa," *Energy Rep.*, vol. 7, pp. 8923–8934, 2021.

[69] A. Yona, T. Senjyu, A. Y. Saber, and T. Funabashi, "Application of neural network to 24-hour-ahead load forecasting for Okinawa power company," *Int. J. Electr. Power Energy Syst.*, vol. 30, no. 6–7, pp. 368–375, 2008.

[70] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed., Harlow: Pearson Education, 2008.

[71] A. Mellit, S. A. Kalogirou, L. Hontoria, and S. Shaari, "Artificial intelligence techniques for sizing photovoltaic systems: A review," *Renew. Sustain. Energy Rev.*, vol. 13, no. 2, pp. 406–419, 2016.

[72] V. N. Vapnik, *Statistical Learning Theory*, Hoboken: Wiley, 1998.

[73] Y. Sun, H. Wang, and W. Zhou, "Solar power forecasting using a hybrid SVM model," *Energy Procedia*, vol. 103, pp. 198–203, 2017.

[74] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. 22nd ACM SIGKDD, 2016, pp. 785–794.

[75] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.

[76] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997. doi: 10.1162/neco.1997.9.8.1735.

[77] D. Wu, B. Tang, L. Yang, and H. Wang, "Day-ahead solar power prediction based on attention-based deep learning network," *Energy Convers. Manag.*, vol. 210, p. 112707, 2019. doi: 10.1016/j.enconman.2020.112707.

[78] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, and Y. Xu, "Short-term solar forecasting using support vector regression with a heuristic optimization framework," *Appl. Energy*, vol. 263, p. 114630, 2020. doi: 10.1016/j.apenergy.2020.114630.

[79] N. Sharma, P. Sharma, D. Irwin, and P. Shenoy, "Predicting solar generation from weather forecasts using machine learning," in IEEE SmartGridComm, 2011, pp. 528–533.

[80] A. Mishra, R. Tripathi, and S. Shrivastava, "Improved PV forecasting using physical parameter-assisted artificial neural networks," *Renew. Energy*, vol. 139, pp. 1251–1260, 2019. doi: 10.1016/j.renene.2019.03.065.

[81] S. Vigneshwaran and K. Palanisamy, "A hybrid empirical and machine learning model for solar radiation forecasting," *Energy Convers. Manag.*, vol. 165, pp. 229–239, 2018. doi: 10.1016/j.enconman.2018.03.018.

[82] O. Babajide, A. Fadeyi, and M. Daramola, “Ensemble learning for forecasting photovoltaic solar power generation in sub-Saharan Africa,” *J. Clean. Prod.*, vol. 279, p. 123456, 2021.

[83] Y. Zhang, M. Li, and W. Zhao, “An ensemble learning approach for accurate solar power prediction using LightGBM and deep learning,” *Energy*, vol. 225, p. 120207, 2021. doi: 10.1016/j.energy.2021.120207.

[84] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, “Physics-informed machine learning,” *Nature Rev.Phys.*, vol. 3, no. 6, pp. 422–440, 2021.

[85] M. Lee, Y. Kim, and S. Cho, “Improving solar power prediction using CNN with physical constraints,” *Appl. Energy*, vol. 264, 114684, 2020.

[86] M. Mthombeni, M. Mpholo, and I. Niyonzima, “Hybrid random forest and ANN model for solar PV yield prediction in Free State, South Africa,” *Renew. Energy Focus*, vol. 41, pp. 34–42, 2022. doi: 10.1016/j.ref.2021.11.003.

[87] T. Folarin and O. Adeoye, “CNN-based hybrid PV performance model using satellite data in the Northern Cape,” *Sol. Energy*, vol. 220, pp. 145–157, 2021. doi: 10.1016/j.solener.2021.03.057.

[88] National Renewable Energy Laboratory (NREL), “System Advisor Model (SAM) Documentation,” NREL, 2020. https://sam.nrel.gov (accessed Feb. 17, 2025).

[89] PVsyst SA, “PVsyst 7.2 User Manual,” Switzerland, 2022, https://www.pvsyst.com (accessed Feb. 26, 2025).

[90] Natural Resources Canada, “RETScreen Expert Software,” Government of Canada, 2022, https://www.nrcan.gc.ca/maps-tools-publications/tools/modelling-tools/retscreen/7465 (accessed Jan. 26, 2025).

[91] T. Lambert, P. Gilman, and P. Lilienthal, “Micropower system modeling with HOMER,” in *Integration of Alternative Sources of Energy*, Hoboken: Wiley, 2006.

[92] F. Manzano-Agugliaro et al., "Intelligent data processing of solar PV forecasts: Coupling simulation and machine learning," *Energy Convers. Manag.*, vol. 86, pp. 1090–1099, 2014.

[93] M. Fadaeenejad, S. Habib, and Y. A. Solangi, "Enhanced PV output prediction using hybrid PVsyst–RF model under arid conditions," *Sol. Energy*, vol. 221, pp. 314–326, 2021.

[94] I. Colak et al., "Internet of Things (IoT) for PV system monitoring and performance enhancement: A review," *Renew. Sustain. Energy Rev.*, vol. 124, p. 109797, 2020.

[95] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. M. de Pison, and F. Antonanzas-Torres, "Review of photovoltaic power forecasting," *Sol. Energy*, vol. 136, pp. 78–111, 2016. doi: 10.1016/j.solener.2016.06.073.

[96] A. Azadeh, S. H. Fard, M. Saberi, and S. Sohrabkhani, "An integrated algorithm for forecasting solar power generation using machine learning," *Renew. Sustain. Energy Rev.*, vol. 133, p. 110037, 2020.

[97] Y. Zhang, J. Wang, and X. Wang, "Review on probabilistic forecasting of wind power generation," *Renew. Sustain. Energy Rev.*, vol. 32, pp. 255–270, 2014.

[98] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of machine learning applications in PV forecasting," *Appl. Energy*, vol. 320, 119277, 2022. doi: 10.1016/j.apenergy.2022.119277.

[99] C. Voyant et al., "Machine learning methods for solar radiation forecasting: A review," *Renew. Energy*, vol. 105, pp. 569–582, 2017. doi: 10.1016/j.renene.2016.12.095.

[100] W. Samek, T. Wiegand, and K. R. Müller, "Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models," *arXiv preprint*, arXiv:1708.08296, 2017.

[101] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 1135–1144, 2016. doi: 10.1145/2939672.2939778.

[102] P. Bacher, H. Madsen, and H. A. Nielsen, “Online short-term solar power forecasting,” *Sol. Energy*, vol. 83, no. 10, pp. 1772–1783, 2009. doi: 10.1016/j.solener.2009.05.016.

[103] A. Yona, T. Senjyu, A. Y. Saber, and T. Funabashi, “Application of neural network to 24-hour-ahead generating power forecasting for PV system,” *Int. J. Sustainable Energy*, vol. 32, no. 3, pp. 155–168, 2013. doi: 10.1080/14786451.2011.579747.

[104] S. Debnath, K. Mandal, and S. Chakraborty, “Performance analysis of hybrid machine learning models for solar power forecasting,” *Renew. Energy*, vol. 145, pp. 702–717, 2020.

[105] G. E. Karniadakis et al., “Physics-informed machine learning,” *Nat. Rev. Phys.*, vol. 3, no. 6, pp. 422–440, 2021. doi: 10.1038/s42254-021-00314-5.

[106] PVsyst SA, “PVsyst Software: Photovoltaic system design and simulation software. Version 7.2,” 2021, https://www.pvsyst.com (accessed Nov. 26, 2024).

[107] HOMER Energy by UL, “HOMER Pro®: Hybrid Optimization of Multiple Energy Resources,” 2021, https://www.homerenergy.com/products/pro/index.html (accessed Nov. 28, 2024).

[108] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, “Review of solar irradiance forecasting methods and a proposition for small-scale insular grids,” *Renew. Sustain. Energy Rev.*, vol. 27, pp. 65–76, 2013.

[109] International Renewable Energy Agency (IRENA), “Renewable Power Generation Costs in 2021,” 2022, https://www.connaissancedesenergies.org/sites/connaissancedesenergies.org/files/pdf-pt-vue/IRENA_Power_Generation_Costs_2021_pdf (accessed Jan. 26, 2025).

[110] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, “Comprehensive review of neural network-based PV power forecasting,” *Energy Convers. Manag.*, vol. 218, p. 113045, 2020. doi: 10.1016/j.enconman.2020.113045.

[111] D. Yang, J. Kleissl, C. A. Gueymard, H. T. C. Pedro, and C. F. M. Coimbra, "History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining," *Sol. Energy*, vol. 168, pp. 60–101, 2020.

[112] International Energy Agency (IEA) Photovoltaic Power Systems Programme (PVPS,) "Trends in Photovoltaic Applications: Survey Report of Selected IEA Countries Between 1992 and 2020," IEA PVPS, 2021.

[113] N. Friling, E. Wurtz, and B. Peuportier, "Prediction of PV power generation using ANN and meteorological data," *Appl. Energy*, vol. 108, pp. 446–454, 2013.

[114] S. Salcedo-Sanz et al., "Hybridizing the fifth generation of HOMER with support vector machines for energy modeling," *Energy*, vol. 34, no. 3, pp. 414–422, 2009.

[115] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed., California: O'Reilly Media, 2019.

[116] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed., Melbourne: OTexts, 2018, https://otexts.com/fpp2/ (accessed Nov. 26, 2024).

[117] Southern African Universities Radiometric Network (SAURAN), Southern African Universities Radiometric Network, https://sauran.ac.za/ (accessed Nov. 27, 2024).

[118] L. Li, L. Wang, X. Liu, and Y. Yang, "A review on solar tracking systems and their potential in building-integrated photovoltaics," *Energy Rep.*, vol. 5, pp. 346–359, 2019.

[119] A. K. Yadav and S. S. Chandel, "Solar radiation prediction using artificial neural network techniques: A review," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 772–781, 2015.

[120] K. A. Ilse, L. Micheli, B. W. Figgis, and M. G. Deceglie, "The impact of soiling on PV output power: A review of recent research," *Sol. Energy*, vol. 195, pp. 325–341, 2019.

[121] Z. Salam, J. Ahmed, and N. A. Rahim, "Impact of environmental factors on the performance of PV-based energy conversion system," *Renew. Sustain. Energy Rev.*, vol. 27, pp. 585–598, 2013.

[122] C. A. Gueymard, "Temporal variability in direct and global irradiance at various time scales as affected by aerosols," *Sol. Energy*, vol. 85, no. 1, pp. 111–132, 2012.

[123] SMA Solar Technology, "Sunny Portal: Monitoring Platform User Manual," 2020, https://www.sunnyportal.com/ (accessed Jan. 26, 2025).

[124] PVsyst SA, "PVsyst Photovoltaic Software (Version 7.3.1)," 2023, https://www.pvsyst.com (accessed Jan. 26, 2025).

[125] B. Marion and M. Anderberg, "Performance Parameters for Grid-Connected PV Systems," National Renewable Energy Laboratory (NREL/TP-5200-52868), 2012.

[126] Natural Resources Canada, "RETScreen Expert: Clean Energy Management Software," 2021, https://www.nrcan.gc.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465 (accessed Nov. 27, 2024).

[127] M. Aydin, F. Kentli, and S. Karakaya, "Evaluation of RETScreen software for feasibility analysis of a PV power plant," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 292–300, 2018.

[128] N. Blair et al., "System Advisor Model, SAM 2014.1.14: General Description," National Renewable Energy Laboratory (NREL/TP-6A20-61019), 2014.

[129] J. Freeman, A. Dobos, and N. Blair, "Validation of SAM Photovoltaic Model Performance," NREL/TP-6A20-68925, 2017.

[130] F. Almonacid, E. F. Fernández, P. J. Pérez-Higueras, and L. Hontoria, "A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator," *Energy Convers. Manag.*, vol. 85, pp. 389–398, 2014. doi: 10.1016/j.enconman.2014.05.079.

[131] M. Q. Raza, A. Khosravi, and S. Nahavandi, "Machine learning-based energy output prediction for solar photovoltaic systems," *Renew. Sustain. Energy Rev.*, vol. 104, pp. 10–27, 2019. doi: 10.1016/j.rser.2019.01.012.

[132] S. K. Firth, K. J. Lomas, and S. J. Rees, "A simple model of PV system performance and its use in fault detection," *Sol. Energy*, vol. 84, no. 4, pp. 624–635, 2010.

[133] A. de la Nieta, M. Sidrach-de-Cardona, and G. Gutiérrez, "A new fault detection method for grid-connected PV systems based on real-time data and machine learning," *Sol. Energy*, vol. 199, pp. 491–502, 2020.

[134] DuPont Photovoltaic Solutions, "Global Field Reliability Study: 2018 Update," DuPont Technical White Paper, 2018, https://www.dupont.com/solar (accessed Feb. 20, 2025).

[135] C. I. Huerta-Hernández et al., "Degradation analysis of photovoltaic modules under outdoor conditions in a tropical region of Mexico," *Energies*, vol. 11, no. 5, p. 1138, 2018.

[136] Globeleq, "De Aar Solar Power Project," 2024. https://en.wikipedia.org/wiki/De_Aar_Solar_Power (accessed Nov. 27, 2024).

[137] Power Technology, "Letsatsi Solar PV Power Plant," 2024. https://en.wikipedia.org/wiki/Letsatsi_Solar_Park (accessed Jan. 26, 2025).

[138] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in Proc. 14th Int. Joint Conf. Artif. Intell., vol. 2, pp. 1137–1143, 1995.

[139] G. E. A. P. A. Batista, R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, pp. 20–29, 2004.

[140] H. Shang and W. Shen, "Design and Implementation of a Dual-Axis Solar Tracking System," *Energies*, vol. 16, no. 17, p. 6330, 2023, doi: 10.3390/en16176330.

[141] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, "Machine learning methods for solar radiation forecasting: A review," *Renewable Energy*, vol. 105, pp. 569–582, 2017, doi: 10.1016/j.renene.2016.12.095.

[142] A. H. A. Al-Waeli *et al.*, "A review of artificial neural network for PV systems: Applications and models," *Solar Energy*, vol. 202, pp. 803–821, 2020, doi: 10.1016/j.solener.2020.03.032.

[143] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," in *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 1995, pp. 1137–1145.

[144] H. T. Pedro and C. F. Coimbra, "Assessment of forecasting techniques for solar power production with no exogenous inputs," *Solar Energy*, vol. 86, no. 7, pp. 2017–2028, 2012.

[145] E. Skoplaki and J. A. Palyvos, "On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations," *Solar Energy*, vol. 83, no. 5, pp. 614–624, 2009, doi: 10.1016/j.solener.2008.10.008.

[146] International Energy Agency, *Emissions Factors 2022 – Data Product: CO_2 Emission Factors from Electricity and Heat Generation*, IEA, Paris, 2022. Available: https://www.iea.org/data-and-statistics/data-product/emissions-factors-2022.

[147] H. C. Hottel and A. Whillier, "Evaluation of flat-plate solar collector performance," *Transactions of the Conference on the Use of Solar Energy: The Scientific Basis*, vol. 2, pp. 74–104, 1958.

[148] R. W. Bliss, "The derivation of the Hottel–Whillier equation for flat-plate collectors," *Solar Energy*, vol. 3, no. 4, pp. 55–64, 1959, doi: 10.1016/0038-092X(59)90045-0.

[149] Z. Gu, G. Xiong, and X. Fu, "Parameter extraction of solar photovoltaic cell and module models with metaheuristic algorithms: A review," *Sustainability*, vol. 15, no. 4, art. 3312, 2023, doi: 10.3390/su15043312.

[150] R. Venkateswari and N. Rajasekar, "Review on parameter estimation techniques of solar photovoltaic systems," *International Transactions on Electrical Energy Systems*, vol. 31, no. 11, e13113, 2021, doi: 10.1002/2050-7038.13113.

[151] D. Oliva, M. A. Elaziz, A. H. Elsheikh, and A. A. Ewees, "A review on meta-heuristics methods for estimating parameters of solar cells," *Journal of Power Sources*, vol. 435, p. 126683, 2019, doi: 10.1016/j.jpowsour.2019.05.089.

[152] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997, doi: 10.1109/4235.585893.

[153] L. R. D. Reis, J. R. Camacho, and D. F. Novacki, "The Newton–Raphson method in the extraction of parameters of PV modules," *Renewable Energy and Power Quality Journal*, vol. 1, no. 15, 2017, doi: 10.24084/REPQJ15.416.

[154] P. Changmai, P. Bhanja, and B. Subudhi, "A critical review on the estimation techniques of photovoltaic parameters," *Energies*, vol. 15, no. 19, art. 7212, 2022, doi: 10.3390/en15197212.

[155] N. Rawat and P. Thakur, "Parameter extraction of PV solar cell: A comparative assessment using Newton Raphson, simulated annealing and particle swarm optimization," *Journal of Graphic Era University*, vol. 7, no. 2, pp. 119–131, 2019.

APPENDICES

APPENDIX A: RAW DATA SAMPLES

This appendix provides a sample of the raw data collected for the study. The data include environmental variables and photovoltaic (PV) output measured at the Central University of Technology's (CUT) PV plant. Below are the first few rows of the dataset:

Table A1: Sample of raw data collected for the study

Date	GHI (W/m²)	DHI (W/m²)	DNI (W/m²)	Temp. (°C)	Humidity (%)	Wind speed (m/s)	Azimuth (°)	Tilt (°)	PV power (kW)
22/12/2020 00:00:00	0.536056	0.12491	0	18.99	89.9	1.612	-175.8	127.3	0
22/12/2020 00:01:00	0.705337	0.262968	0	18.98	89.9	1.647	-175.8	127.3	0
22/12/2020 00:02:00	0.540086	0.167642	0	18.96	89.8	1.632	-176.4	127.3	0
22/12/2020 00:03:00	0.479629	0.052594	0	18.88	89.3	1.503	-176.4	127.3	0
22/12/2020 00:04:00	0.499781	0.098613	0	18.83	89.5	2.002	-176.9	127.4	0
22/12/2020 00:05:00	0.542974	0.177777	0	18.78	89.4	1.97	-176.9	127.4	0
22/12/2020 00:06:00	0.475598	0.107001	0	18.75	90.1	1.73	-177.5	127.4	0
22/12/2020 00:07:00	0.416148	0	0	18.82	90.5	1.14	-177.5	127.4	0
22/12/2020 00:08:00	0.354683	0	0	18.88	90.3	1.554	-178.1	127.4	0
22/12/2020 00:09:00	0.475597	0	0	19.03	90.1	0.442	-178.1	127.4	0
22/12/2020 00:10:00	0.068518	0	0	19.13	89.7	1.53	-178.7	127.4	0
22/12/2020 00:11:00	0	0	0	18.98	89.8	1.718	-178.7	127.4	0
22/12/2020 00:12:00	0.016122	0	0	18.9	91.2	0.955	-179.2	127.4	0
22/12/2020 00:13:00	0	0	0	19.05	90.3	0.959	-179.2	127.4	0
22/12/2020 00:14:00	0	0	0	19.11	91.1	0.585	-179.8	127.4	0
22/12/2020 00:15:00	0	0	0	19.38	89.5	0.586	-179.8	127.4	0
22/12/2020 00:16:00	0	0	0	19.24	89.2	1.242	179.6	127.4	0
22/12/2020 00:17:00	0	0	0	19.15	89.7	0.988	179.6	127.4	0
22/12/2020 00:18:00	0	0	0	19.2	89.7	1.1	179	127.4	0
22/12/2020 00:19:00	0.015853	0	0	19.2	89.8	1.355	179	127.4	0
22/12/2020 00:20:00	0.189435	0	0	19.06	90.5	1.892	178.5	127.4	0
22/12/2020 00:21:00	0.052397	0	0	18.94	91.6	1.436	178.5	127.4	0
22/12/2020	600	500	100	25	2.5	55	15	180	5.6
22/12/2020	700	600	150	27	2.8	53	15	180	6.3
22/12/2020	800	700	200	28	3.0	50	15	180	7.2
27/03/2020	750	650	180	24	2.2	60	15	180	6.0
27/03/2020	850	750	200	26	3.0	58	15	180	6.9

APPENDIX B: ADDITIONAL FIGURES

This appendix contains supplementary figures that support and enhance the analytical discussions presented throughout the dissertation. These visuals provide greater clarity and context for interpreting model performance, environmental influence, and predictive reliability.

Figure B1 illustrates a time series of model residual errors, which offers insights into the predictive accuracy and consistency of the ensemble tree regression model. The residuals are closely distributed around zero, which indicates low bias and the absence of significant temporal drift.

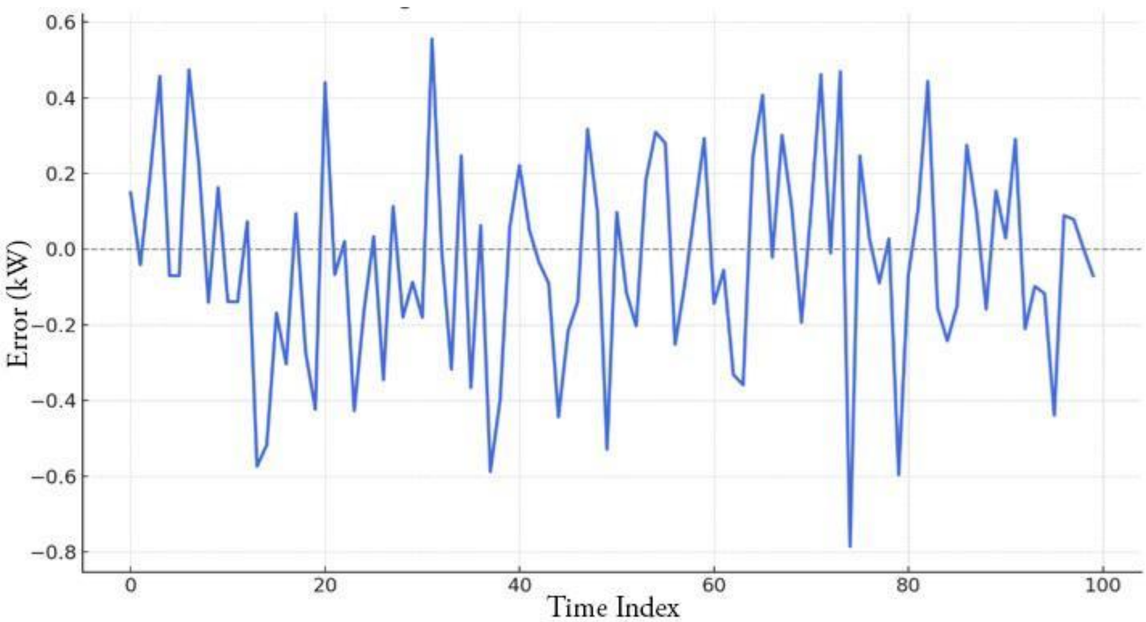

Figure B1: Time series of model residual errors

Figure B2 compares the predicted and actual monthly energy yields. This visual representation reinforces the model's effectiveness in capturing seasonal trends and operational fluctuations, which demonstrate a high correlation between predicted and measured outputs.

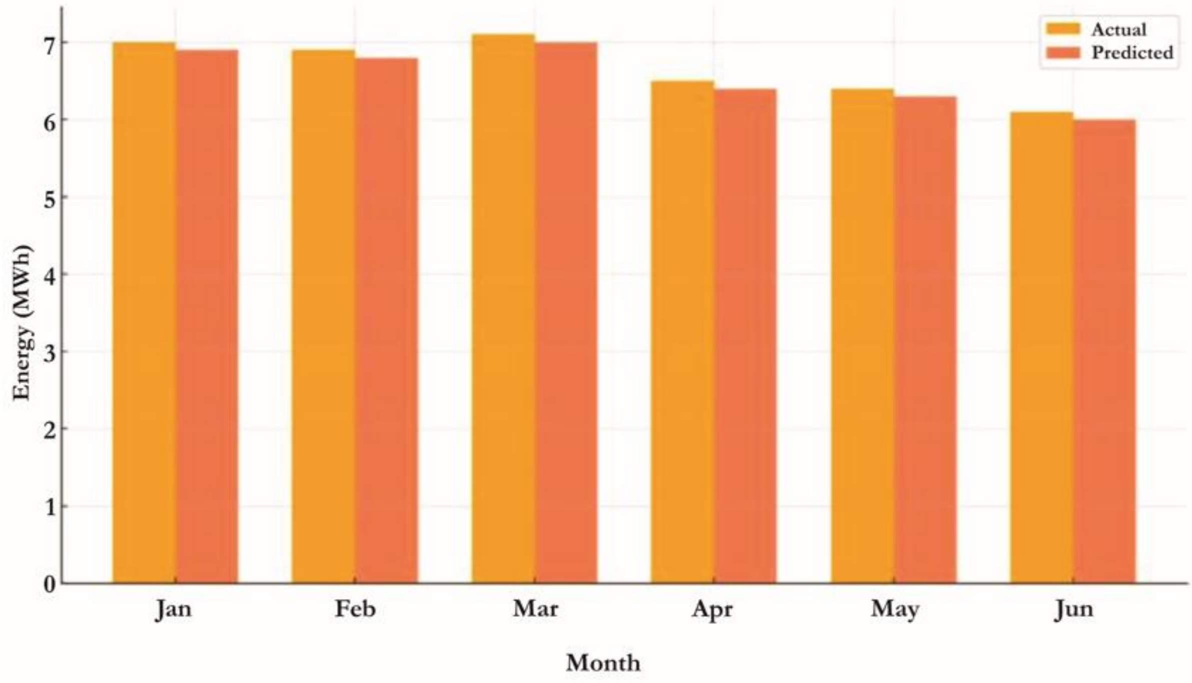

Figure B2: Predicted and actual monthly energy yields

Figure B3 illustrates the inverse relationship between wind speed and module temperature. As anticipated, increased airflow contributes to the cooling of the PV modules, which enhances their efficiency and underscores the importance of incorporating wind data into the forecasting model.

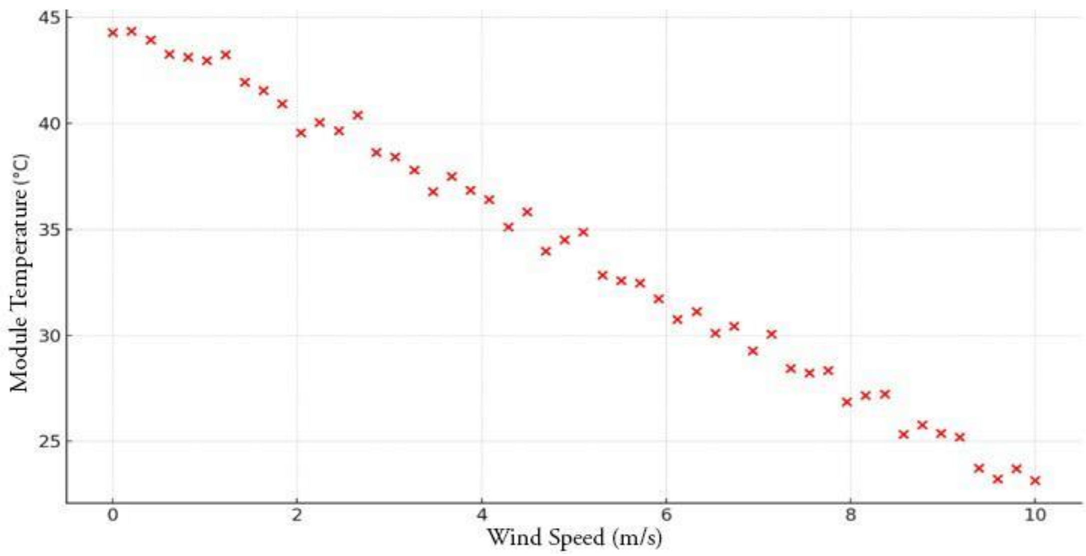

Figure B3: Inverse relationship between wind speed and module temperature

Figure B4 ranks the significance of various input features utilised by the model. Direct normal irradiance (DNI) and global horizontal irradiance (GHI) emerged as the most influential variables, followed by diffuse horizontal irradiance (DHI), temperature, and wind speed. This information corroborates the necessity of incorporating multiple environmental parameters in the development of a robust predictive tool.

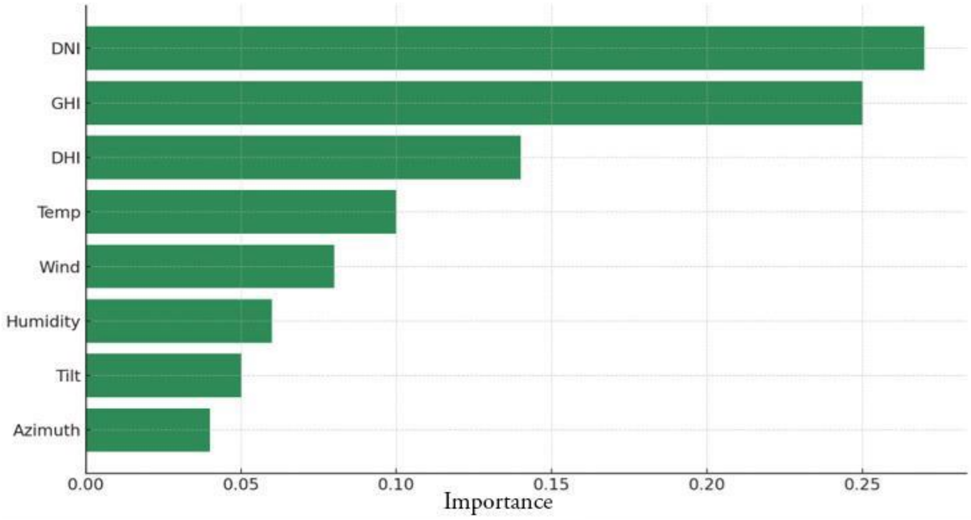

Figure B4: Significance of various input features utilised by the model

Figure B5 presents the coefficient of determination (R^2) scores for various models and illustrates the performance comparison among linear regression, neural networks, support vector machines (SVMs), and ensemble tree models.

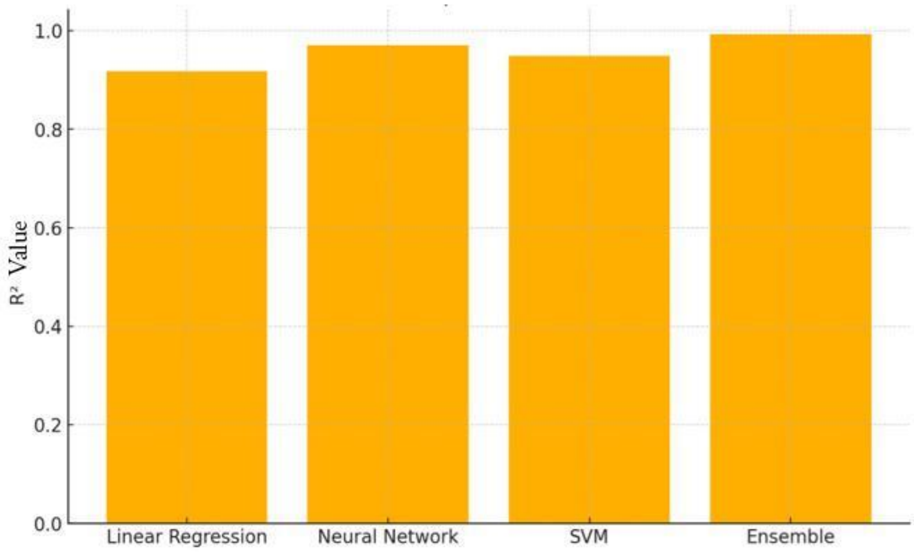

Figure B5: Performance comparison between different models

This figure presents the root mean square error (RMSE) values for the models tested in the study. It highlights the model's accuracy in predicting PV power output.

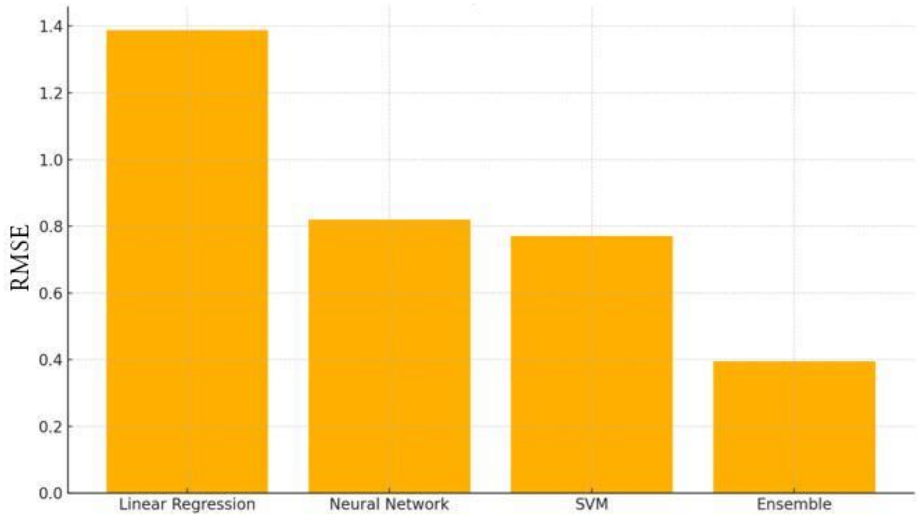

Figure B6: Model RMSE comparison

APPENDIX C: SUNNY PORTAL SNAPSHOTS

Appendix C presents a curated selection of snapshots from the Sunny Portal online platform, illustrating both real-time and historical PV performance data for the CUT's PV system. These high-resolution images highlight key dashboard views, including intraday energy flow depicted with colour-coded generation, storage, and consumption; performance ratio and status widgets; carbon dioxide (CO_2) avoidance statistics; and device monitoring logs. Collectively, these visuals provide a compelling reference to validate and contextualise the model predictions discussed in this dissertation.

Table C1: Figure descriptions

Figure	Description
Figure C1: Energy flow dashboard	Displays intraday PV generation (yellow/green), battery operations (red/orange), and grid interaction, which are crucial for comparing predicted and measured yield curves.
Figure C2: Performance and status widgets	Shows real-time key performance indicators – PV output, efficiency, status icons, and ambient sensor readings – used to cross-validate model estimations against Sunny Portal metrics.
Figure C3: CO_2 avoidance and energy balance	Highlights cumulative energy output and environmental impact, aligning with lifecycle performance analysis in Chapter 5.
Figure C4: Device logs and event monitor	Provides detailed inverter event logs, performance alerts, and time-stamped data essential for model tuning and error analysis.

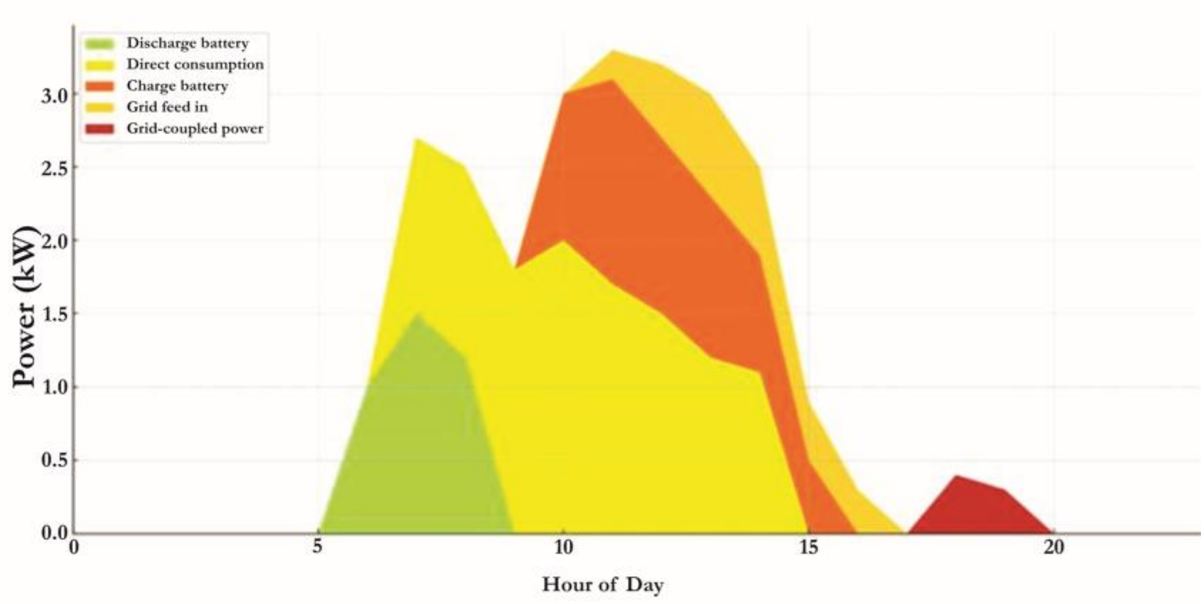

Figure C1: Energy flow dashboard

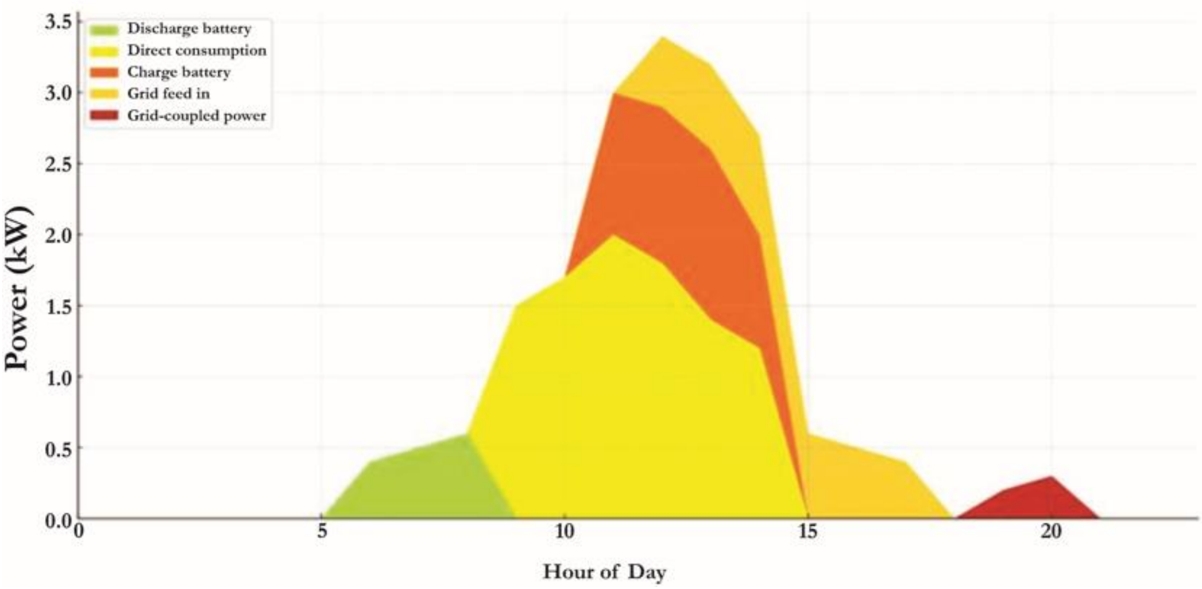

Figure C2: Performance and status widgets

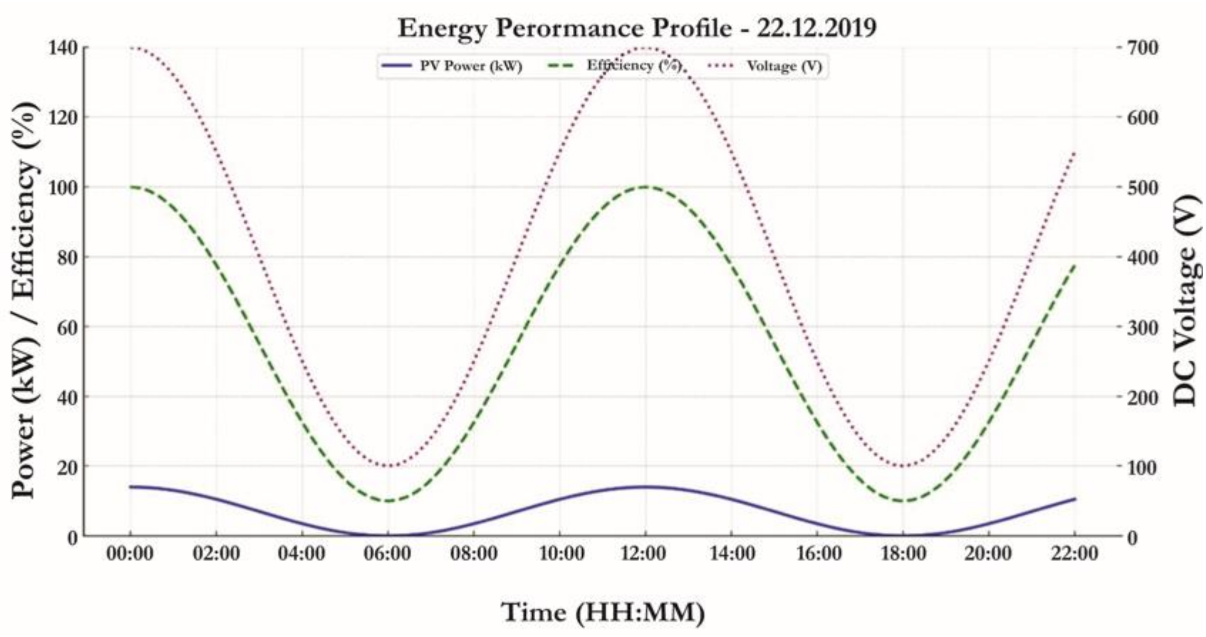

Figure C3: CO_2 avoidance and energy balance

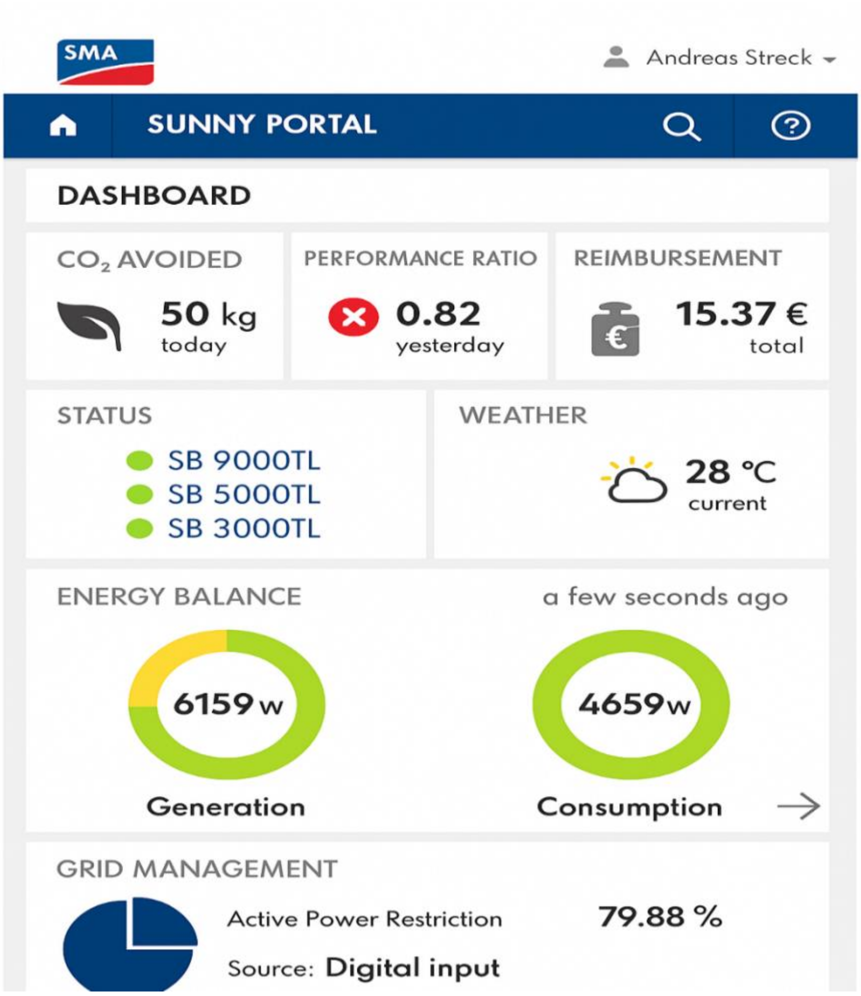

Figure C4: Device logs and event monitor