A Bayesian Approach To Maximise Photovoltaic System Output

Masters Dissertation

School of Computer Science & Applied Mathematics
University of the Witwatersrand

Keanu Noel
1913371

Supervised by Professor Ritesh Ajoodha

May 26, 2025

A dissertation submitted to the Faculty of Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Masters of Science by dissertation

Abstract

This thesis addresses the critical issue of optimising photovoltaic (PV) system output, an essential objective in the pursuit of efficient and scalable renewable energy solutions. As global energy demands rise and concerns over climate change intensify, solar power has emerged as a leading solution for sustainable electricity generation. However, the performance of PV systems is highly sensitive to environmental factors which can vary significantly across seasons and geographical locations. These fluctuations create a complex optimisation problem in determining the most effective system configuration that can dynamically adapt to seasonal and regional variations in solar potential. Traditional approaches often rely on fixed or rule-based models that do not adequately account for these variations, leading to suboptimal energy yields and the inefficient use of solar infrastructure.

In this research, a Bayesian Network model is developed to learn the conditional dependencies between meteorological variables (such as solar irradiance, temperature, and wind speed) and PV system configuration parameters (tilt angle, orientation, inverter properties). By using a Bayesian approach, the developed model accommodates uncertainty and dynamically adjusts PV system tilt configurations to weather variations, aiming to maximise PV output. Score-based methods are employed to construct the network structure, and Maximum Likelihood Estimation (MLE) to determine the Conditional Probability Distributions (CPDs) of the network. Additionally, Maximum a Posteriori (MAP) estimation is applied to identify the optimal seasonal PV system tilt configurations in light of specific weather conditions.

Key findings demonstrate the effectiveness of the model in optimising PV output by offering adaptive configuration strategies that respond to local seasonal meteorological patterns. This includes the superior performance of the Hill Climb Search algorithm compared to Simulated Annealing for structure learning, the utilisation of MAP Estimation for identifying optimal PV system tilt configurations under varying meteorological conditions, and the statistically significant advantage of dynamic configurations over fixed installations for enhancing PV system output. These results underscore the potential of Bayesian approaches for data-driven optimisation in renewable energy systems.

This research provides a robust framework that enhances PV system performance and contributes to the growing body of knowledge on renewable energy optimisation through probabilistic modelling. Ultimately, this research presents a novel, data-driven methodology which informs the design and operation of more efficient PV systems, answering both the “so what” and “now what” in the context of sustainable energy advancements.

Key Concepts: Bayesian Network, Structure Learning, MAP Estimation, Photovoltaic Systems, Optimisation

Acknowledgments

I would like to express my deepest gratitude to my family and friends for their unwavering support throughout this journey. To my Mom, Sylvia, and my Dad, Dion, your encouragement, sacrifices, and belief in me have been my greatest motivation. To my brother, Warren, and my sister, Clio, your constant support and understanding have been invaluable. This work would not have been possible without your love and guidance. Thank you for always being there for me.

I would also like to express my sincere gratitude to my supervisor, Prof. Ritesh Ajoodha, for the continuous support of my dissertation, for his patience, motivation, and immense knowledge.

Declaration

I, Keanu Noel, hereby declare that the contents of this Masters Dissertation are my own work. This dissertation is being submitted for the degree of Master of Science by dissertation in Computer Science at the University of the Witwatersrand. This work has not been submitted to another university or for any other degree.

Keanu Noel
Student Number: 1913371
May 26, 2025

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Solar Energy Efficiency in the Face of Global Energy and Climate Challenges

In an era characterised by rapid climate change and a growing global energy crisis, solar energy has emerged as a cornerstone for sustainable development. With growing concerns over climate change, energy security, and air pollution, countries around the world are increasingly turning to solar energy as a key component of their energy portfolios. As the globe grapples with the pressing need to shift away from traditional energy sources, the appeal of harnessing solar power grows stronger than ever. Governments, organisations, and households are progressively investing in photovoltaic (PV) systems, not only to reduce carbon footprints, but also as a strategic step towards energy independence and long-term cost savings [International Energy Agency 2023]. However, as the adoption of solar power increases, so does the need to optimise these systems for maximum efficiency. An inefficient solar array not only reduces the return on investment, but it also wastes potential environmental benefits. With billions of capital being invested in solar infrastructure around the world, it is crucial to ensure that these systems run optimally.

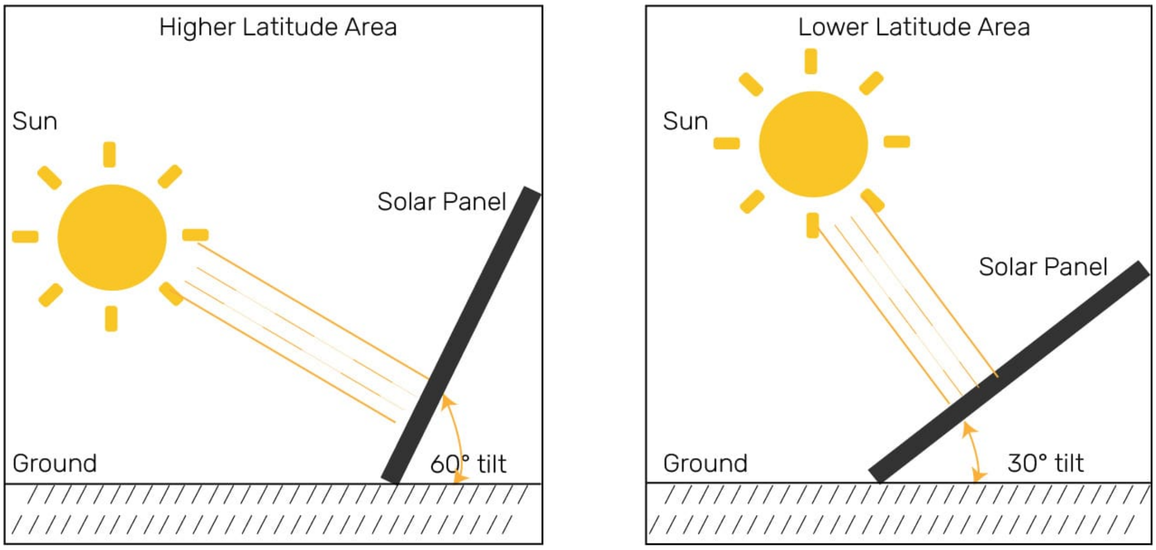

Figure 1.1: Accurately Configuring the Tilt of a PV System Based on the Geographical Latitude of a Location Assists with Maximum Solar Absorption.

As the demand for sustainable energy solutions grows, maximising the efficiency of photovoltaic (PV) systems has become increasingly critical. However, optimising PV system output remains challenging due to the complex interplay between meteorological conditions and system configuration variables. Figure 1.1 depicts a portion of this complexity by illustrating how the latitude of a location impacts the optimised orientation of a system's solar panels. Generally, higher latitudes require larger tilt angles to better capture the lower sun angles typical of those regions, while lower latitudes, closer to the equator require smaller tilts due to the sun's higher position in the sky throughout the year. Traditional static optimisation models are limited by their inability to adapt to dynamic environmental conditions, which results in suboptimal PV performance. This research addresses this problem by developing a Bayesian Network model that probabilistically accounts for meteorological variability and determines the optimal seasonal configurations to maximise PV output.

This research utilises the strengths of Explainable AI (Artificial Intelligence) to identify the optimal tilt configuration of PV systems in order to maximise PV output under fluctuating weather conditions. This study provides a robust framework for optimising solar energy systems by investigating novel AI approaches. This study provides the potential to increase the financial viability of solar investments while also contributing to the larger objective of a sustainable and greener future.

The methodology employed within this study includes acquiring historic meteorological and PV system data as well as utilising modern Machine Learning approaches to optimise and evaluate PV system configurations. Machine Learning (ML) techniques provide a powerful framework for representing probabilistic dependencies amongst variables in a system and offer a flexible approach to modeling the dynamic nature of a PV system. Moreover, their probabilistic capabilities allow for the incorporation of diverse sources of information, including meteorological data, historical energy production, and system parameters, thus enhancing the accuracy and robustness of results.

1.2 Research Questions

This research addresses the problem of maximising the energy output of a photovoltaic system by investigating several critical research questions. A central focus is placed on understanding and modelling the complex, probabilistic relationships between meteorological variables and PV system parameters. The study questions how these interdependencies can be effectively captured using data-driven approaches, particularly within a Bayesian framework, to support informed configuration decisions. An essential aspect of this investigation involves evaluating the accuracy and reliability of the derived optimal PV system tilt configurations. Furthermore, the research explores the impact of employing dynamic, seasonally adjusted PV system tilt configurations as opposed to traditional fixed-tilt setups. By quantifying the performance differences between these strategies, the study provides empirical evidence on whether adaptive tilt adjustment offers significant improvements in energy output, thereby contributing to more efficient PV system design and operation.

1.3 Methodology Overview

This study employs Bayesian Networks to learn the complex dependencies between meteorological and PV system variables. The methodology enables the optimal PV system tilt configurations to be learnt, thereby providing maximum PV output on a seasonal basis.

Modern data-driven approaches are employed to obtain high-resolution meteorological data and to simulate the performance of various photovoltaic (PV) system tilt configurations under realistic environmental conditions. The meteorological data includes solar irradiance, temperature, wind speed, and other relevant atmospheric variables that influence PV output. By integrating and pre-processing this data, a comprehensive and structured dataset is developed, serving as the foundation for modeling and analysis. To uncover the underlying probabilistic dependencies between meteorological factors and PV system performance, structure learning techniques are applied within the Bayesian Network framework. Specifically, two score-based structure learning algorithms are explored: Simulated Annealing and the Hill Climb Search algorithm with strategic enhancements such as the incorporation of Tabu Lists and Random Restarts. These enhancements promote a more thorough exploration of the solution space and increase the likelihood of identifying a globally optimal network structure. Once the network structure is established, Maximum A Posteriori (MAP) estimation is utilised to infer the most optimal PV system tilt configurations which result in the maximum PV output, given observed meteorological conditions. This inference process allows for data-driven identification of optimal system parameters under varying environmental conditions.

To evaluate the effectiveness of these configurations, a robust experimental design is constructed. This involves simulating PV output under both fixed and dynamic tilt settings, with particular attention given to seasonal variability. By comparing system performance across these configurations, the study quantitatively assesses the potential gains in energy output associated with adaptive, seasonally optimised tilt strategies.

This study contributes a novel framework that integrates many sources of uncertainty and dependencies in PV system performance, resulting in more efficient optimisation compared to traditional methods. The approaches within this study strive to provide an adaptable and efficient PV system by ensuring PV system configurations are optimised for a variety of meteorological conditions, resulting in improved overall performance and reliability.

1.4 Dissertation Overview

This thesis is structured to provide a logical progression from the theoretical foundations of photovoltaic system optimisation to the application and evaluation of the proposed methodology. The discussion begins in Chapter 2, which offers a critical review of related work in the field. This chapter examines both traditional optimisation techniques and more recent probabilistic and Bayesian approaches applied to PV system performance enhancement. Through this review, gaps in the current literature are identified, providing the motivation and context for the present study. Chapter 3 introduces the methodological framework adopted in this research. It details the process followed to optimise seasonal PV system tilt configurations, including data acquisition, preprocessing, simulation of PV output, and the application of Bayesian Network-based techniques for structure learning and inference. The chapter outlines how these methods are used to model the probabilistic relationships between meteorological variables and PV system parameters, with the aim of identifying optimal tilt configurations.

Chapters 4, 5, and 6 present the core results of the study and offer an in-depth discussion of their implications. These chapters explore the effectiveness of the proposed optimisation approach, assess the comparative performance of fixed versus dynamic tilt strategies, and evaluates the accuracy of the inferred configurations under varying environmental conditions. The analysis is supported by statistical validation and visualisation to ensure a rigorous interpretation of the results. The final chapter, concludes the thesis by summarising the key findings and contributions of the research. It reflects on the limitations encountered and discusses the broader implications of the study for PV system design and operation. The chapter also outlines directions for future research.

Chapter 2

Related Work

Several studies have investigated optimising PV systems to increase output efficiency. While some studies focus on traditional optimisation techniques, others have employed Bayesian methods to investigate aspects of this problem. The following chapter examines existing literature, with the aim of identifying key methodologies, challenges, and opportunities in this prominent domain.

2.1 Conceptual Framework

A PV system consists of solar modules, also known as solar panels, that include several photovoltaic cells composed of semiconductor materials such as silicon. When sunlight strikes these cells, it produces direct current (DC) power. This electricity is then passed via an inverter, which converts it to alternating current (AC), the form of electricity most typically utilised in homes and businesses. Additional components that aid optimum performance and safety may be included in the system, such as mounting structures, charge controllers, batteries (in off-grid or hybrid systems), and monitoring equipment [Kalogirou 2014]. The output of a photovoltaic system is commonly defined as the quantity of electrical energy generated during a specific time period, expressed in kilowatt-hours (kWh). This output is determined by a number of elements, including solar irradiation, panel orientation, ambient temperature, shading conditions, and overall system efficiency. The tilt of the solar modules is particularly important, as it determines the angle at which sunlight strikes the panel surface. The output can be measured at the DC stage or after AC conversion, however AC output is usually utilised for practical performance study because it reflects the actual consumable energy [Masters 2013].

PV systems are grouped into traditional grid-connected, off-grid and hybrid systems. Grid-connected systems are directly connected to the public electrical grid and are primarily intended to supplement energy supply by exporting excess power back to the grid. Off-grid systems run independently of the grid and often rely on battery storage to provide a constant energy source. Hybrid systems integrate PV systems with other energy sources, such as diesel engines or wind turbines and provide greater reliability in environments with fluctuating supply or high energy needs [Kalogirou 2014].

Kalogirou [2014] and Masters [2013] suggest that the effectiveness of PV system output is inextricably linked to the choice of a PV system parameters, specifically the tilt configuration as well as the variability of meteorological conditions. Traditional PV systems with fixed tilt configurations often fail to capture the maximum potential energy output when faced with fluctuating weather patterns. Duffie and Beckman [2013] note that this has resulted in a greater emphasis placed on the value of dynamic PV system tilt configurations, that can adjust to changing environmental circumstances. Implementing adaptive techniques is critical for increasing PV system efficiency and dependability, especially in unexpected and diverse climates. The advantage of employing dynamic PV systems over fixed configurations is depicted in Figure 2.1, which illustrates a simulated scenario where dynamic systems consistently achieve higher monthly energy outputs.

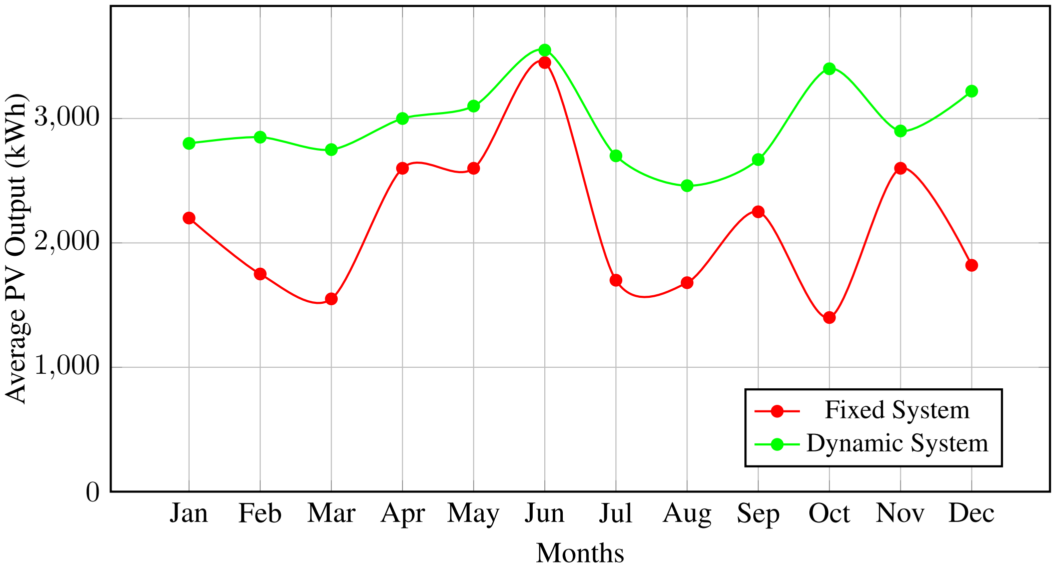

Figure 2.1: Dynamic configurations outperform fixed installations.

Vasisht *et al.* [2016], Keyrouz [2018], Abdelrahman *et al.* [2016], and Ajmal *et al.* [2020] have all contributed to a better understanding of how to optimise PV systems under changing environmental conditions, particularly through the use of adaptive and dynamic configuration strategies. This shared focus underscores the critical role of optimisation in enhancing the efficiency and reliability of PV systems.

Ajmal *et al.* [2020] specifically address the issue of partial shading in PV arrays, which can significantly reduce energy output. Their study analyses static and dynamic reconfiguration strategies, demonstrating that dynamic approaches outperform static setups. Static reconfiguration distributes shading effects uniformly by using pre-fixed architectures such as series-parallel (SP) or total-cross-tied (TCT) structures, while dynamic reconfiguration entails real-time modifications utilising algorithms or switching matrices, allowing the system to respond to changing shading conditions throughout the day to optimise power production.The dynamic reconfiguration reduces the negative impacts of partial shading by constantly optimising the array architecture to maintain higher energy production.

Vasisht *et al.* [2016] provides the groundwork by investigating how seasonal variables, such as changes in solar irradiance and temperature, affect the performance of PV installations. Their findings show the intrinsic unpredictability in PV production, emphasising the importance of dynamic adaptation to these oscillations to achieve optimal performance throughout the year.

The superior performance of dynamic PV systems compared to fixed can be seen as a reoccurring theme in studies by Ajmal *et al.* [2020] and Vasisht *et al.* [2016]. The research conducted emphasises the importance of dynamic and adaptive systems that can adjust to changing environmental conditions and maintain optimal system performance.

Both Keyrouz [2018] and Abdelrahman *et al.* [2016] employ Bayesian techniques for enhancing Maximum Power Point Tracking (MPPT) controllers. The methods use Bayesian Estimation to improve the tracking of the maximum power point under changing environmental variables such as shade or irradiance. These methods are recognised for their capacity to successfully manage the uncertainty and variability inherent in solar energy systems, resulting in a more adaptable and accurate control mechanism compared to traditional approaches. The adoption of Bayesian approaches demonstrates a shared recognition of the importance of probabilistic models in optimising PV system efficiency, particularly in dynamic and uncertain contexts.

Integrating these insights offers a compelling rationale for pursuing dynamic configurations over fixed installations in PV systems. Vasisht *et al.* [2016] highlights environmental variability as a necessary condition for adaptive methods. Keyrouz [2018] demonstrates that Bayesian-based optimised MPPT controllers are effective at dynamically adapting to these environmental changes. Finally, Ajmal *et al.* [2020] demonstrates that dynamic reconfiguration, particularly in reaction to partial shading, improves PV system performance, compared to a fixed configuration.

These studies collectively support the hypothesis that dynamic configurations are more effective than fixed configurations in maximising PV output. By dynamically adjusting to environmental conditions and mitigating external factors such as partial shading, these configurations ensure more consistent and higher energy production, aligning with the overall goal of optimising PV system performance across varied and unpredictable environments. This rationale forms the basis for further exploration and validation of dynamic tilt strategies in traditional PV system design and management.

2.2 Traditional Approaches for PV System Optimisation

The optimisation of PV systems has gained an increased amount of attention in recent years due to the growing demand for efficient and dependable renewable energy solutions. Researchers such as Sangeetha *et al.* [2024] and Khatib and Muhsen [2020] have offered detailed overviews of the challenges and advancements in this field, focussing on critical aspects such as system sizing control and the application of intelligent optimisation approaches. One of the most prevalent approaches described in these studies is the application of metaheuristic algorithms. These techniques, which include the Genetic Algorithm (GA), Particle Swarm Optimisation (PSO), and Differential Evolution (DE), have shown great promise in solving complex optimisation problems in PV systems. They are specifically employed to improve system performance under practical restrictions such as shading, temperature variability, and changing irradiance levels.

2.2.1 Metaheuristic Algorithms

Genetic Algorithm (GA)

Metaheuristic algorithms are widely utilised for optimising the configuration and performance of PV systems. The primary objective in this context is to maximise the power output (P_{out}) by tuning essential system parameters, such as the tilt angle (θ), inverter efficiency (η), irradiance (G) and temperature (T). The Genetic Algorithm (GA) is one such method that seeks to identify the optimal PV system configuration by iteratively maximising P_{out}, expressed as a function of these parameters [Goldberg 1989]. The fitness function, which evaluates the quality of a given solution, can be represented as:

$$f(\mathbf{x}) = P_{\text{out}}(\theta, \eta, G, T),$$

where $\mathbf{x} = (\theta, \eta, G, T)$ represents the decision variables.

The GA process begins with a selection phase, where high-performing solutions, referred to as parents, are chosen from the current population. The selection is guided by the fitness values of the solutions, ensuring that those with higher performance are more likely to influence subsequent generations. Commonly used methods for selection include roulette-wheel and tournament selection, which prioritise diversity while retaining strong candidates. Following selection, the algorithm proceeds to the crossover phase. Here, the attributes of the selected parent solutions are combined to create offspring configurations. For instance, given two parent solutions $\mathbf{x}_1$ and $\mathbf{x}_2$, the offspring solution is formed as a weighted combination of the parents, as expressed by:

$$\mathbf{x}_{\text{offspring}} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2,$$

where α is a parameter that controls the contribution of each parent to the offspring. This step ensures that promising traits from both parents are inherited by the offspring, promoting convergence toward optimal configurations. To enhance exploration and prevent premature convergence to local optima, the mutation phase introduces random perturbations into the offspring solutions. This stochastic modification expands the search space and allows the algorithm to discover potentially superior configurations that may have been overlooked during the crossover phase. By iteratively applying selection, crossover, and mutation, the Genetic Algorithm refines the population of solutions over successive generations. This approach enables GA to effectively balance exploration and exploitation, making it a robust and versatile tool for PV system optimisation.

Particle Swarm Optimisation (PSO)

Particle Swarm Optimisation is a stochastic, population-based optimisation technique initially introduced by Kennedy and Eberhart [1995]. It is inspired by the social dynamics observed in natural swarms, such as bird flocking or fish schooling. In PSO, a group of particles (candidate solutions) traverse the search space by iteratively adjusting their positions and velocities based on both their individual experiences and the collective experience of the swarm. Each particle is influenced by its personal best-known position and the global best-known position found by the entire swarm, thereby facilitating convergence toward optimal or near-optimal solutions in complex and high-dimensional spaces [Clerc and Kennedy 2002; Poli *et al.* 2007]. Due to its conceptual simplicity, computational efficiency, and adaptability, PSO has been widely adopted in solving engineering optimisation problems, including renewable energy system design and control [Eberhart and Shi 2001]. For each particle i, its position $\mathbf{x}_i = (\theta_i, \eta_i, G_i, T_i)$ and velocity $\mathbf{v}_i$ are updated based on personal and global best positions:

$$\mathbf{v}_i(t+1) = \omega\mathbf{v}_i(t) + c_1r_1(\mathbf{p}_i - \mathbf{x}_i(t)) + c_2r_2(\mathbf{g} - \mathbf{x}_i(t)),$$

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1),$$

where $\mathbf{p}_i$ is the personal best position of particle i and is the global best position found by the swarm. The inertia weight is represented by ω, controlling the influence of the previous velocity, c_1 and c_2 are acceleration coefficients that balance exploration and exploitation and r_1, r_2 are random numbers uniformly distributed in $[0, 1]$, ensuring stochastic movement.

Differential Evolution (DE)

Differential Evolution is an established metaheuristic algorithm extensively employed in PV system optimisation due to its simplicity, robustness, and effectiveness in handling non-linear and non-differentiable functions [Storn and Price 1997]. Unlike many other optimisation algorithms, DE leverages the concept of differential mutation and applies vector-based operations to iteratively refine candidate solutions within a population. In the context of PV system optimisation, DE seeks to maximise power output (P_{out}) by adjusting critical system parameters, including the tilt angle (θ), inverter efficiency (η), irradiance (G) and temperature (T). Each candidate solution within the population is represented as a vector of decision variables:

$$\mathbf{x}_i = (\theta_i, \eta_i, G_i, T_i), \quad i = 1, 2, \ldots, N_p,$$

where N_p denotes the population size. The optimisation process begins with the mutation phase, in which a mutant vector is generated for each candidate solution. This vector is derived by combining the difference between two randomly chosen population vectors and adding the result to a third vector. Formally, the mutant vector can be expressed as:

$$\mathbf{v}_i = \mathbf{x}_{r1} + F(\mathbf{x}_{r2} - \mathbf{x}_{r3}),$$

where $\mathbf{x}_{r1}, \mathbf{x}_{r2}, \mathbf{x}_{r3}$ are randomly selected vectors from the population, $F \in [0, 2]$ is the scaling factor that determines the amplification of the difference vector, and $\mathbf{v}_i$ represents the resulting mutant vector. Next, a crossover operation is performed to enhance diversity within the population. The mutant vector is combined with the original target vector to produce a trial vector, where individual components of the mutant vector are retained based on a predefined crossover probability (C_r). At least one component from the mutant vector is always included, ensuring that the trial vector incorporates new information. This process is described mathematically as:

$$u_{i,j} = \begin{cases} v_{i,j}, & \text{if } \text{rand}_j \leq C_r \text{ or } j = j_{\text{rand}}, \\ x_{i,j}, & \text{otherwise.} \end{cases}$$

Finally, the selection phase determines whether the trial vector or the original target vector will advance to the next generation. This decision is based on the fitness function $f(\mathbf{x}) = P_{\text{out}}(\theta, \eta, G, T)$, with the vector achieving the higher fitness value being selected. Formally, this can be expressed as:

$$\mathbf{x}_i^{\text{new}} = \begin{cases} \mathbf{u}_i, & \text{if } f(\mathbf{u}_i) > f(\mathbf{x}_i), \\ \mathbf{x}_i, & \text{otherwise.} \end{cases}$$

The iterative application of metaheursitc algorithms effectively balance exploration and exploitation, gradually converging towards optimal configurations. Their ability to navigate complex solution spaces and accommodate diverse constraints makes this domain of algorithms a particularly valuable tool in PV system optimisation. In addressing challenges such as environmental variability and system constraints, metaheursitc algorithms have demonstrated reliable identification of configurations that maximise P_{out}, further solidifying their role as a robust optimisation technique in PV systems [Sangeetha *et al.* 2024].

2.2.2 Metaheuristic Approaches

Differential Evolution for Optimal Sizing of Standalone PV Systems

The optimisation of standalone photovoltaic (PV) systems involves striking a balance between reliability and cost-effectiveness, typically measured through two primary objectives: minimising the Loss of Load Probability (LLP) and the Life Cycle Cost (LCC) [Khatib and Muhsen 2020; Storn and Price 1997]. LLP quantifies the likelihood that the system fails to meet load demands, calculated as the ratio of unserved load to total load demand:

$$\text{LLP}(\mathbf{x}) = \frac{\sum(\text{unserved load})}{\sum(\text{load demand})}.$$

LCC represents the total cost of the PV system over its operational lifetime, encompassing installation, maintenance, and replacement costs. These objectives are governed by the decision variables $\mathbf{x} = (N_{\text{PV}}, N_{\text{bat}})$, where N_{PV} denotes the number of PV panels and N_{bat} represents the battery capacity.

Differential Evolution is particularly well-suited for addressing such multi-objective optimisation problems due to its robustness and simplicity. In this application, DE evolves a population of candidate solutions through the iterative processes of mutation, crossover, and selection. Mutation introduces diversity by generating mutant vectors that explore the search space, while crossover blends information from mutant and target vectors to create trial solutions. Selection ensures that only solutions with improved fitness (measured as a trade-off between LLP and LCC) are retained for subsequent generations. By iteratively refining the population, DE converges toward a set of Pareto-optimal solutions, enabling system designers to evaluate trade-offs between reliability and cost. This method has proven effective in identifying configurations that minimise LLP while keeping LCC within acceptable limits, offering a practical and reliable approach for standalone PV system sizing in diverse operational contexts.

Metaheuristic algorithms possess the ability to improve power output and efficiency, but lack standardised benchmarking, limiting direct performance comparison [Sharma *et al.* 2023]. By leveraging a metaheuristic approach to optimise initial costs and complexity associated with sizing standalone PV systems, Khatib and Muhsen [2020] demonstrates that the AI optimisation can significantly speed up the sizing process while maintaining a relatively high accuracy, aligning with Sangeetha *et al.* [2024] around the emphasis on the role of AI in PV system optimisation. The application of AI is also linked to a broader trend of incorporating intelligent algorithms to deal with the complexity and nonlinearity of PV systems. Additionally, while the algorithms' effectiveness is noted, further details on empirical results or precise improvements in system outputs could have strengthened their findings. The article also suggests AI-based algorithms, hybrid techniques, and improved sensor technologies can considerably increase PV system efficiency and efficacy, particularly in adverse environmental conditions.

Intelligent Optimisation of PV Inverter Systems

The integration of AI-based algorithms is explored further by Zhang *et al.* [2024], who investigates the control and intelligent optimisation of PV inverter systems. Their analysis aims to enhance stability and control performance by minimising control error and power fluctuations while ensuring adaptability to external grid variations. This involved addressing a multi-objective optimisation problem defined as:

$$\min\left(\text{Control Error}(\mathbf{x}), \text{Power Fluctuation}(\mathbf{x})\right), \tag{2.1}$$

where $\mathbf{x} = (\theta_{\text{inverter}}, \phi_{\text{control}}, \gamma_{\text{stability}})$ encompasses key control parameters such as the inverter angle (θ_{inverter}), control phase (ϕ_{control}), and stability factor ($\gamma_{\text{stability}}$).

Zhang *et al.* [2024] emphasises that traditional control approaches are losing effectiveness as PV systems increase and integrate into larger power networks. To overcome these problems, they argue for the use of intelligent control approaches, which are intended to improve system stability and performance, despite their inherent complexities. This perspective complements the findings of Khatib and Muhsen [2020], who also identify the limitations of conventional approaches and the potential of AI to revolutionise PV system optimisation.

2.3 Disadvantages of Traditional Optimisation Strategies

While Sangeetha *et al.* [2024], Khatib and Muhsen [2020], and Zhang *et al.* [2024] provide valuable insights into the optimisation of photovoltaic (PV) systems using a variety of advanced techniques, including AI-driven sizing algorithms, intelligent control mechanisms, and heuristic-based performance tuning, a recurring limitation becomes apparent across these studies. Despite the sophistication of these methods, they often rely on deterministic or data-intensive models that are not always well-suited for the complex, uncertain environments in which PV systems operate.

Traditional AI-based optimisation and control strategies, such as artificial neural networks (ANNs), fuzzy logic controllers, and genetic algorithms, are generally dependent on extensive historical datasets to train and validate models. This reliance introduces several challenges. Firstly, these methods typically assume a stationary environment, meaning that the statistical properties of inputs (e.g., solar irradiance, temperature, wind speed) do not change over time. In reality, however, PV system performance is strongly influenced by non-stationary and stochastic factors, such as rapidly shifting weather patterns, seasonal solar angle variability, and unforeseen system degradation or maintenance issues.

Secondly, many of these models are black-box in nature, meaning they can offer predictions or decisions without providing interpretable justifications. This opacity poses a problem in critical engineering contexts where understanding model behaviour and uncertainty is vital. Furthermore, the optimisation process itself in AI and heuristic methods can be computationally intensive and prone to overfitting, particularly when the solution space is high-dimensional and noisy, as is often the case in PV configuration problems.

Given these challenges, it becomes evident that while traditional AI-based and heuristic optimisation techniques have advanced the field of PV system performance enhancement, they remain limited in their ability to effectively manage the inherent uncertainty and variability of real-world conditions. Their dependence on static models, opaque decision-making processes, and high data or computational demands can hinder adaptability and long-term reliability. These limitations highlight the need for more flexible, interpretable, and uncertainty-aware frameworks. In this context, probabilistic approaches, particularly those grounded in Bayesian reasoning offer a promising avenue for more robust and informed optimisation of PV systems under real-world complexity.

2.4 Bayesian Optimisation in PV Systems

Bayesian techniques offer versatile and efficient solutions in dealing with the inherent uncertainties and complexities of solar energy systems. The diverse research conducted demonstrates the broad application of Bayesian approaches in PV systems. Chiodo *et al.* [2015] focus on the availability analysis of PV inverters, a critical component for system dependability. In contrast, Ciaramella *et al.* [2016] and Hu *et al.* [2023] focus on power forecasting and MPPT, with a clear emphasis on energy output optimisation. Leicester *et al.* [2016] employ a different approach, utilising Bayesian Networks to assess the economic viability of PV systems, confirming the method's relevance in financial decision making.

Despite the diverse application of research, all four studies utilise a Bayesian framework to address various aspects of PV system performance, emphasising its strengths in probabilistic modelling and uncertainty management. For example, Chiodo *et al.* [2015] use Bayesian analysis to assess the availability of PV inverters in uncertain conditions, whereas Ciaramella *et al.* [2016] and Hu *et al.* [2023] use a Bayesian Neural Network model to forecast solar power output, with a focus on predictive accuracy and confidence interval estimation.

Ciaramella *et al.* [2016] provides a Bayesian Neural Network model that pairs probabilistic inference with neural network capabilities. This method captures both the intrinsic nonlinearity of the PV power output and the uncertainty in the forecasting process. The Bayesian framework is useful because it provides a probabilistic estimation of the forecast, thus providing a measure of confidence in the predictions. The model leverages meteorological and historical PV power data to make predictions. Key meteorological variables such as temperature, solar irradiance, and cloud cover are used as inputs to forecast future PV power output. Bayesian Networks enable the incorporation of prior knowledge and handle uncertainty more effectively than traditional neural networks. This approach also allows the model to produce probabilistic outputs, which is useful for understanding the range and reliability of predictions. The results show that the Bayesian-based Neural Network model performs effectively, with improved accuracy over traditional methods, particularly in capturing fluctuations in PV power output.

The utilisation of Bayesian techniques combined with neural networks, illustrates the trend of hybrid Bayesian machine learning to improve a PV system's adaptability. Studies by Ciaramella *et al.* [2016] and Hu *et al.* [2023] highlight the ability of neural networks to capture complicated, non-linear correlations in data, while the Bayesian approach adds a layer of uncertainty quantification.

Bayesian learning provides an effective method for learning optimal photovoltaic system configurations that maximise output by incorporating both prior knowledge and empirical data. These techniques excel at explaining the relationships and dependencies between variables and their effect on system performance. Unlike traditional ML techniques, models such as Bayesian networks update dynamically as new data is integrated, providing a more nuanced understanding of how uncertainty affects PV production. This adaptability not only improves the precision of accurate system designs, but also provides a clearer picture of the uncertainties involved, resulting in better informed and robust decision-making in PV system design and management.

2.4.1 Bayesian Neural Network Framework

The following framework covers the fundamental mathematical structure of Bayesian Neural Networks as utilised in the above studies for PV power forecasting, from the specification of priors and likelihood, to posterior approximation and prediction.

Bayesian Inference and Prior Distribution on Weights

In a Bayesian Neural Network, we place a prior distribution on each weight in the network. Let $\mathbf{w}$ represent the set of all weights in the network. A common choice is a Gaussian prior [Neal 1995]:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w} \mid 0, \sigma^2). \tag{2.2}$$

Likelihood of the Observed Data

Given a set of input-output pairs $\{\mathbf{x}_i, y_i\}_{i=1}^N$ where y_i is the observed PV power output for input $\mathbf{x}_i$, the likelihood of observing data D given the weights $\mathbf{w}$ is typically assumed to be Gaussian [Bishop 2006]:

$$p(D \mid \mathbf{w}) = \prod_{i=1}^{N} p(y_i \mid \mathbf{x}_i, \mathbf{w}) = \prod_{i=1}^{N} \mathcal{N}(y_i \mid f(\mathbf{x}_i; \mathbf{w}), \sigma_y^2), \tag{2.3}$$

where $f(\mathbf{x}_i; \mathbf{w})$ is the output of the neural network for input $\mathbf{x}_i$ given weights $\mathbf{w}$.

Posterior Distribution of Weights

Using Bayes' theorem, it is possible to calculate the posterior distribution of weights given data D:

$$p(\mathbf{w} \mid D) = \frac{p(D \mid \mathbf{w})\, p(\mathbf{w})}{p(D)}, \tag{2.4}$$

where $p(D) = \int p(D \mid \mathbf{w})\, p(\mathbf{w})\, d\mathbf{w}$ is the marginal likelihood [Murphy 2012].

Approximation of the Posterior using Variational Inference

Since the posterior $p(\mathbf{w} \mid D)$ is intractable, variational inference (VI) can be used to approximate it. In VI, we define a variational distribution $q(\mathbf{w} \mid \boldsymbol{\theta})$ with parameters $\boldsymbol{\theta}$ and minimise the Kullback-Leibler (KL) divergence between $q(\mathbf{w} \mid \boldsymbol{\theta})$ and the true posterior $p(\mathbf{w} \mid D)$ [Blei *et al.* 2017]:

$$\mathrm{KL}(q(\mathbf{w} \mid \boldsymbol{\theta}) \parallel p(\mathbf{w} \mid D)) = \int q(\mathbf{w} \mid \boldsymbol{\theta}) \log \frac{q(\mathbf{w} \mid \boldsymbol{\theta})}{p(\mathbf{w} \mid D)}\, d\mathbf{w}. \tag{2.5}$$

Evidence Lower Bound (ELBO)

Minimising the KL divergence is equivalent to maximising the evidence lower bound (ELBO), which is used as the objective function in variational inference [Jordan *et al.* 1999]:

$$\mathcal{L}(\boldsymbol{\theta}) = \mathbb{E}_{q(\mathbf{w}|\boldsymbol{\theta})} [\log p(D \mid \mathbf{w})] - \mathrm{KL}(q(\mathbf{w} \mid \boldsymbol{\theta}) \parallel p(\mathbf{w})). \tag{2.6}$$

Prediction with Bayesian Neural Networks

After training, predictions are acquired by integrating over the posterior distribution of weights for new input $\mathbf{x}^*$:

$$p(y^* \mid \mathbf{x}^*, D) = \int p(y^* \mid \mathbf{x}^*, \mathbf{w})\, p(\mathbf{w} \mid D)\, d\mathbf{w}. \tag{2.7}$$

This integral can be approximated by sampling weights from the posterior distribution $p(\mathbf{w} \mid D)$ and averaging the predictions [Blundell *et al.* 2015]. A point prediction can be obtained by taking the expectation of $p(y^* \mid \mathbf{x}^*, D)$, which yields the mean of the predictive distribution:

$$\hat{y}^* = \mathbb{E}_{p(y^*|\mathbf{x}^*, D)}[y^*]. \tag{2.8}$$

2.5 Significance of Bayesian Optimisation in PV System Configurations

Bayesian Networks (BNs) offer a considerable advantage over traditional methods for optimising PV System designs due to their intrinsic explainability. In contrast to black-box models, BNs provide a visible framework for understanding the probabilistic correlations between variables, which is critical in complex systems like PV configurations where many variables such as solar intensity, temperature, and equipment efficiency interact. This transparency not only enhances knowledge of the model's decision-making process, but also allows for easier troubleshooting and system refinement. Furthermore, the explainability of BNs enables stakeholders to validate the model against domain knowledge, ensuring that the system's behaviour is consistent with realistic expectations and scientific principles [Derks and De Waal 2020].

In addition to their explainability, BNs excel at handling uncertainty and learning relationships within data, which are critical in the dynamic environment of PV systems. They can incorporate prior information and new evidence to update beliefs, enabling for more informed decision-making in the face of uncertainty [Heckerman 1998]. This ability to model and reason under ambiguity is especially useful in PV systems, where conditions are continuously changing and accurate forecasts are difficult. By understanding the causal relationships and dependencies between numerous variabes, BNs can optimise PV system configurations to maximise efficiency and reliability, resulting in higher energy production and investment returns [Leicester *et al.* 2016]. The BNs' capacity to handle uncertainty and learn from data helps to create robust and adaptive PV systems that can tolerate the fluctuation inherent in renewable energy sources.

2.6 Conclusion

This chapter highlights the evolution of optimisation strategies for PV systems, from traditional approaches to more advanced Bayesian optimisation techniques. Traditional methods, while promising, often fall short in managing the complexity and uncertainty inherent in PV systems. They typically rely on deterministic models that do not account for the probabilistic nature of solar energy generation, leading to less robust system configurations. BN's model the stochastic behavior of PV systems, offering a more nuanced understanding of the interplay between various factors. The strength of BNs lies in their ability to provide a clear and explainable model structure, which is essential for stakeholders to trust and act upon the optimisation results. Furthermore, BNs are adept at handling uncertainty and learning from data, which enables them to adapt to new information and improve over time. The study leverages the strengths of Bayesian Networks, providing a novel approach to identify optimal PV system tilt configurations. It improves the efficiency and reliability of PV systems while also providing a framework that is both transparent and responsive to varying scenarios. This research unleashes the full potential of solar energy, paving the path for a more sustainable and resilient green future.

Chapter 3

Methodology

3.1 Introduction

Optimising the output of a photovoltaic system (PV) is pivotal in improving the efficiency and viability of solar energy as a sustainable energy source. The performance of a PV system is influenced by a multitude of factors, including meteorological variables (such as solar irradiance, temperature, wind speed, and humidity) and PV system parameters (such as panel orientation and tilt angle). Accurately modeling and optimising these variables to maximise the energy output presents a complex, multidimensional problem.

3.2 Problem Statement

The problem of optimising PV system output can be formulated as a mathematical optimisation problem where the objective is to maximise the energy output W of the PV system. The probabilistic energy output W is dependent on various meteorological variables $\hat{E}$ and PV system variables $\hat{X}$, where $\hat{E}$ represents all meteorological variables such that $\hat{E}$ = $\{E_1$, E_2, ... $E_n\}$ and $\hat{X}$ represents all PV System variables such that $\hat{X}$ = $\{X_1$, X_2, ... $X_n\}$.

The objective is to find the optimal set of PV system variables X_{MAP} which maximise the energy output W under given meteorological conditions $\hat{E}$:

$$X_{\mathrm{MAP}} = \arg\max_{\hat{X}} \left(\log P(\hat{E}|\hat{X}) + \log P(\hat{X}) \right), \tag{3.1}$$

where taking the logarithm is done for numerical stability and ease of computation.

3.3 Research Questions and Experimental Setup

To address this optimisation problem, Bayesian Networks (BNs) are utilised. BNs offer a robust framework to model the dependencies and interactions between variables, incorporating both the inherent uncertainties in meteorological data and the non-linear relationships among the variables affecting PV system output. By leveraging BNs, it is possible to capture the probabilistic nature of the variables and to derive an optimal configuration of PV system parameters which maximise the expected PV output.

The methodology addresses the following research questions:

1. **Which structure learning procedure optimises the likelihood of PV system output given the dependencies between the meteorological and PV system variables?**

   This research employs two structure learning methodologies namely, the Hill Climb Search and Simulated Annealing. Each candidate structure's Conditional Probability Distributions (CDPs) is learnt using two techniques, namely Maximum Likelihood Estimation and Bayesian Estimation. The ground truth data distribution, $\mathcal{D}$, is partitioned into training, validation, and test sets. The training time for each technique is recorded to monitor and compare their efficiency in finding the optimal network structure using the training and validation sets. The predictive accuracy of each network is evaluated using standard evaluation metrics with K-Fold Cross Validation. The candidate structure with the highest overall scores is selected as the optimal structure G.

2. **What is the most likely seasonal configuration of PV system variables which maximises PV system output?**

   Using the optimal network G, MAP estimation is employed to solve for the optimal seasonal PV system configurations, given the meteorological variables. The MAP estimations' accuracy in capturing the maximum posterior probability is determined by forward sampling the optimal Bayesian Network. By visually plotting the PV system output provided by the samples and the MAP estimates, the accuracy and reliability of the MAP estimates are examined to ensure they at least provide the highest local PV system output under different meteorological conditions. These techniques facilitate a thorough analysis of the posterior distribution, enhancing the reliability of the MAP estimates in optimising PV system configurations to maximise PV system output.

3. **What is the impact of implementing optimal PV system configurations per season, compared to maintaining a fixed configuration throughout the year?**

   This research question aims to investigate whether employing season-specific optimal photovoltaic (PV) system configurations leads to a statistically significant improvement in PV output compared to maintaining a fixed configuration throughout the year. The study makes use of the optimal Bayesian Network G and MAP estimation to identify optimal configurations for each season, given the meteorological data, followed by a robust experimental design and statistical analysis to evaluate the performance differences between the two systems. The goal of the statistical experiment will aim to highlight the benefit of dynamically optimising PV system configurations compared to maintaining fixed configurations, thereby providing a clearer understanding of the positive impact on PV system output.

3.4 Methodology

The following section illustrates the three phases of the methodology utilised to examine and address the research questions. It provides a detailed explanation of each phase, outlining the specific techniques and approaches employed. This structured methodology ensures a comprehensive analysis and facilitates a systematic approach to deriving meaningful insights.

3.4.1 Phase 1: Data Collection, Exploratory Data Analysis and Engineering

Publicly available meteorological data is sourced using the Photovoltaic Geographical Information System portal (PVGIS), an extremely reliable and reputable source. PVGIS provides an online API allowing access to solar irradiation and weather data for specific geographic locations at different temporal resolutions.

Using the API, meteorological data is collected for three cities in South Africa (Johannesburg, Cape Town and Durban). These cities were selected by making use of a database containing South Africa's most populated cities [Simplemaps 2024].

Table 3.1: Meteorological Data Attributes

Variable	**Unit**
Location	-
Season	-
Latitude	°
Longitude	°
Date	-
Global Plane of Array Irradiance	W/m²
Direct Plane of Array Irradiance	W/m²
Diffuse Plane of Array Irradiance	W/m²
Temperature	°C
Wind Speed	m/s
Tilt Angle	°
Panel Type	-
Panel Area	m²
Panel Material	-
Cells in Series	-
Parallel Strings	-
Short Circuit Current	A
Open Circuit Voltage	V
Max Power Point Current	A
Max Power Point Voltage	V
Inverter Type	-
AC Voltage	V
Standby Losses	W
Max AC Output Power	W
Peak DC Input Power	W
Peak DC Voltage	V
Max DC Voltage	V
Max DC Current	A
MPPT Voltage Range (Low)	V
MPPT Voltage Range (High)	V
CEC Inverter Type	-
AC Output	W

Creating a Synthetic Photovoltaic System

PVLib is an open-source Python package used to simulate and optimise the performance of photovoltaic (PV) energy installations. It includes tools for simulating solar irradiance, PV system components, and overall performance. This research leverages PVlib's built-in modules (solar panel configurations) and inverters, which are precise representations of real-world components based on manufacturer specifications, which are also validated by independent testing. The systems are part of the databases maintained by the US National Renewable Energy Laboratory (a premier laboratory in the United States dedicated to research, development, commercialisation, and deployment of renewable energy and energy efficiency technologies), which ensures that PVlib simulations use correct and reliable data. This allows for precise modelling of PV system performance, which aids users in building and optimising solar energy installations for research purposes.

The original meteorological dataset provides input for the simulated systems. The dataset is thereafter overlaid with the PV system output of these systems, thus providing data containing PV system output, using different system configurations. The final dataset used for modelling includes a wide range of variables relevant to photovoltaic system performance, covering both meteorological conditions and PV system configuration parameters. A detailed overview of these attributes is provided in Table 3.1.

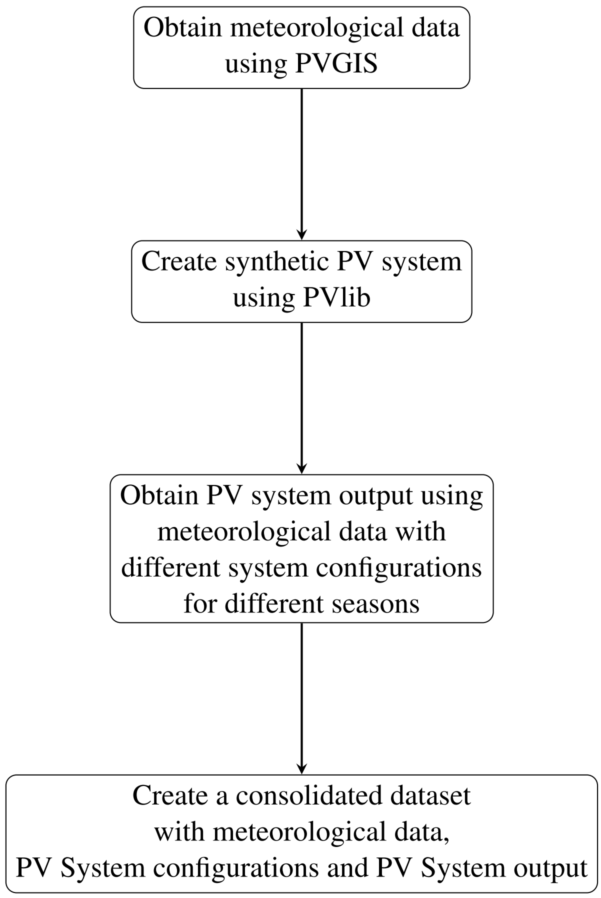

Figure 3.1: Flow of Dataset Creation

Discretisation of Continuous Variables

Continuous variables are frequently discretised in Bayesian Network (BN) modelling to facilitate the representation and computation of probabilistic relationships. The process entails transforming continuous variables into a finite number of discrete intervals or categories [Nojavan *et al.* 2016]. The primary reasons for discretisation include:

- Simplification of Conditional Probability Tables (CPTs)
- Reduction of Computational Complexity
- Improved Interpretability
- Compatibility with Existing Bayesian Tooling Methods

For the purpose of this research, the binning of the meteorological and AC Output variables is done using domain knowledge. Since the main objective is to understand the statistical distributions of the photovoltaic (PV) system that influence PV output, domain knowledge guides the selection of these bin ranges. This approach ensures that the most relevant aspects of the data are captured, while avoiding over complicating the model with unnecessary divisions.

Using the Freedman-Diaconis Rule for Tilt Bins

The Freedman-Diaconis Rule is employed to calculate the optimal bin width and the number of bins for discretising the tilt field. This rule is widely recognised for balancing the trade-off between variability in the data and the size of the dataset. It computes the bin width h using the following formula:

$$h = 2 \cdot \frac{\text{IQR}}{n^{1/3}},$$

where IQR is the inter-quartile range ($Q3 - Q1$), representing the spread of the middle 50% of the data, and n is the total number of observations [Freedman and Diaconis 1981]. Using this formula, the tilt field data points are divided into 113 bins. This discretisation ensures that the bin sizes reflects the natural variability of the data while avoiding excessive granularity.

Importance of the Freedman-Diaconis Rule

1. Data Representation and Bias Reduction

Choosing appropriate bins is critical when discretising continuous variables, especially for probabilistic models like Bayesian Networks. Improperly selected bin widths can introduce bias. Narrow bins may cause overfitting or instability, while wide bins may obscure meaningful patterns. The Freedman-Diaconis Rule strikes a balance, enabling an accurate representation of the tilt field's distribution.

2. Impact on Bayesian Network Structure

The discretised bins of the tilt field play a vital role in constructing the Bayesian Network:

- **Conditional Probability Tables (CPTs):** The bins define the discrete states of the tilt variable in the network's CPTs. Accurate binning ensures that these states capture the true variability of the data, leading to better modeling of dependencies between variables.
- **Inference Accuracy:** The Bayesian Network relies on discretised variables to perform inference. Well-selected bins enhance the network's ability to make precise predictions by aligning the discretisation with the data's natural structure.

A detailed account of the tilt bins can be referenced in Appendix A. The meteorological and AC Output bins are illustrated in Tables 3.2, 3.3 and 3.4.

Bin Range	**Label**
$-\infty$ to 0	Freezing
0 to 10	Cold
10 to 20	Mild
20 to 30	Warm
30 to $+\infty$	Hot

Table 3.2: Custom Bins and Labels for Temperature

Bin Range (m/s)	**Label**
0 to 1.5	Calm
1.5 to 3.3	Light Air
3.3 to 5.5	Light Breeze
5.5 to 7.9	Gentle Breeze
7.9 to 10.7	Moderate Breeze
10.7 to 13.8	Fresh Breeze
13.8 to 17.1	Strong Breeze
17.1 to $+\infty$	High Wind

Table 3.3: Custom Bins and Labels for Wind Speed Based on the Beaufort Scale

Bin Range	**Label**
$-\infty$ to 0	No Output
0 to 100	Low Output
100 to 500	Moderate Output
500 to 1000	High Output
1000 to $+\infty$	Very High Output

Table 3.4: Custom Bins and Labels for AC Output

Distributions of Data

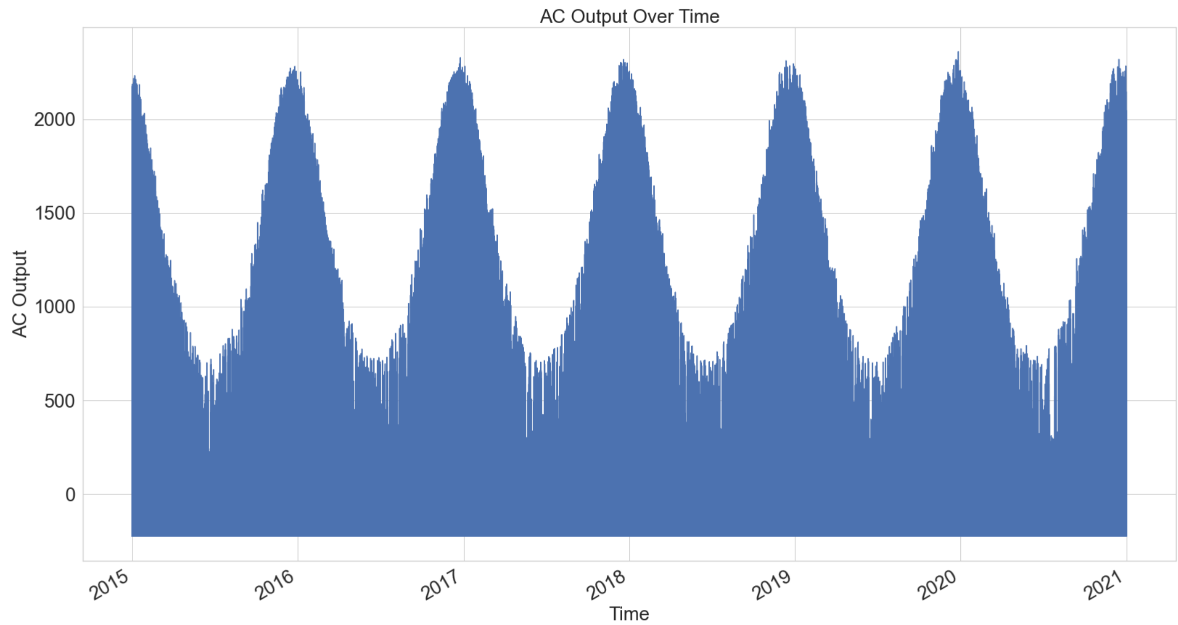

Figure 3.2: AC Output Over Time

Figure 3.2 illustrates the seasonality within the dataset, by depicting maximum PV output during the Summer season in South Africa, with troughs occurring during Winter.

Panel Name 1	Panel Name 2
AstroPower APX 120 (2001)	Kyocera Solar PV75 (2003)
Schott Solar ASE 300 (1999)	BP Solar SX365 (2007)
Shell Solar SQ80 (2003)	Shell Solar SP150 (2001)
BP Solar BP4170 (2003)	Shell Solar SR100 (1999)
Kyocera Solar KC167G (2003)	Siemens Solar SP130 (2001)
Photowatt PW1000 (2000)	Kyocera Solar KC120 (1999)
Solarex MSX 120 (2000)	Shell Solar SP140 (2002)
Sharp ND 123U1 (2003)	First Solar FS 45 (2003)
Yingli Solar YL220 (2008)	SunPower SPR 220 (2008)
Sanyo HIP 210NKHE1 (2008)	AstroPower APX 90 (1999)
Sharp NT R5E1U (2002)	Shell Solar SP150 (2002)
Solarex MST 43LV (1999)	BP Solar BP3150 (2003)
Siemens Solar ST10 (1999)	Sanyo HIP 220HDE1 (2008)
Uni Solar PVL 87 (2003)	Kyocera Solar KC150 (2002)
SunPower SPR 230 (2007)	Sharp NE Q5E2U (2002)

Table 3.5: List of Solar Panels Used in Study

Inverter Name 1	Inverter Name 2
SMA SB7700TL 208V	Growatt 12000 TL3 277V
Growatt 3600 MTL 240V	Power-One PVI 3.6 240V
ABB PVI 5000 277V	Solectria PVI 85 208
Motech PVMate 3000U 240V	ABB UNO 2.0 TL 277V
Satcon AE 30 60 240V	CPS SC20KTL 480V
Solectria PVI 5000S 240V	Power Electronics FS0750CU
KACO blueplanet 1502x 240V	ABB PVI 6000 208V
Fronius CL 60 277V	SMA SB8000TL 240V
Solectria SGI 500XTM 380V	Power-One PVI 6000 208V
Schneider Conext XC540 NA	ABB UNO 7.6 TL 277V
Solectria PVI 13 kW 208V	Schneider Conext CL 18000NA
PV Powered PVP75KW 480V	AEG Protect MPV 480V
PV Powered PVP1100EVR 120V	ABB PVI 3.8 240V
KACO blueplanet 5.0 TL 240V	Chint Solar CHPI20KTL
Schneider GT30 208V	Sinexcel Rainbow 12.5

Table 3.6: List of Inverters

The selection of a diverse set of photovoltaic (PV) panels and inverters utilised in the study, as detailed in Tables 3.5 and 3.6, reflects the variability observed in real-world PV system deployments. In practice, individuals and institutions employ a wide range of PV technologies based on factors such as cost, availability, brand preference, and system requirements. To ensure that the findings of this study are broadly applicable and representative of realistic conditions, a heterogeneous mix of PV components was intentionally incorporated. This approach enables the simulation, modelling and optimisation of system performance across different configurations, thereby enhancing the generalisability and practical relevance of the results.

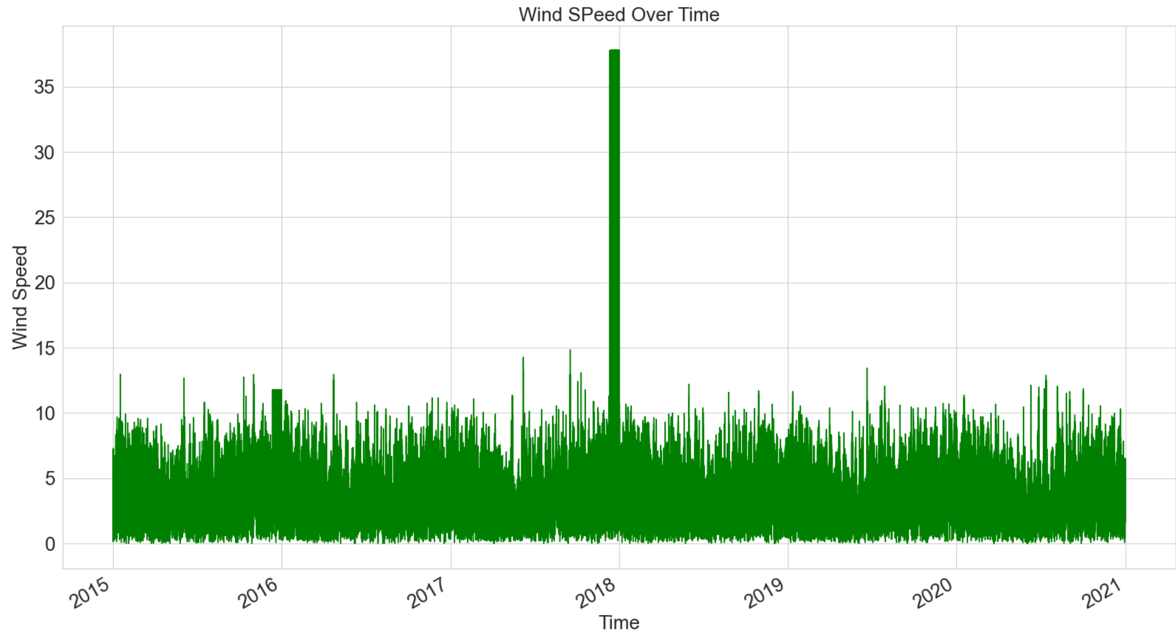

Figure 3.3: Wind Speed Over Time

Figure 3.3 depicts stable wind speeds, with an anomaly detetcted during 2018. It is possible that South Africa may have experienced extreme weather conditions during this period.

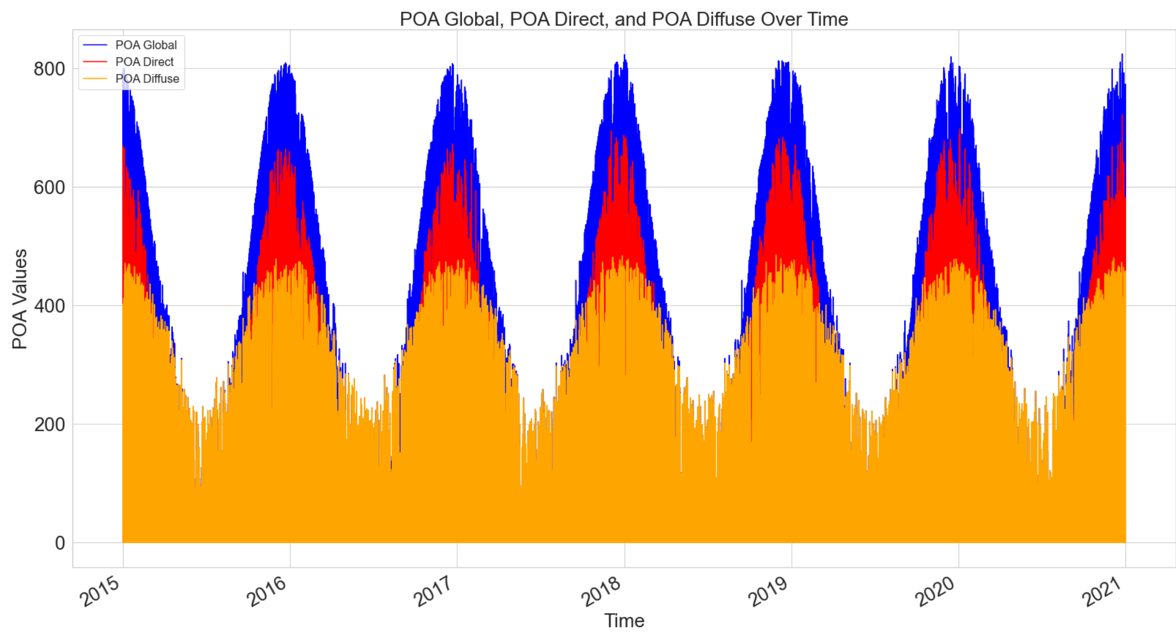

Figure 3.4: POA Over Time

Similar to Figure 3.2, Figure 3.4 illustrates common trends.

3.4.2 Phase 2: Model Development

Structure Learning

Learning the structure of a BN is essential as it outlines the dependencies and relationships that exist between variables. The structure provides a graphical representation of how multiple variables interact and influence one another, which is critical for effective PV system configuration management and accurately maximising PV system output. An accurate representation of relationships within a BN makes the model more interpretable. This transparency is vital for residential and commercial entities who need to understand the reasons behind certain relationships to make informed decisions regarding their PV system configuration [Shin *et al.* 2020]. Unlike complex machine learning models, BNs provide insights into the causal mechanisms at play, which can be crucial for detecting issues and improving operational strategies.

Learning the optimal structure of a BN is an NP-hard task [Koller and Friedman 2009], therefore the time to solve it grows exponentially with the size of the input data. This computational intractability derives from the process of searching through an exponential number of alternative network configurations [Neapolitan 2004]. Therefore, this research will explore utilising existing efficient score based network learning algorithms and heuristics to approximate the optimal structure. This domain of structure learning algorithms aims to find the graph structure that maximises a scoring function based on the data. The scoring function evaluates the goodness of fit of a graph structure in relation to the data, typically penalising complex models to avoid overfitting. The study explores two score-based structure learning algorithms: Hill Climb Search (HCS) and Simulated Annealing (SA).

Hill Climb Search (HCS) is a greedy local search algorithm that iteratively modifies the BN structure by adding, deleting, or reversing edges, accepting only changes that improve a scoring metric such as the Bayesian Information Criterion (BIC) or Bayesian Dirichlet equivalent uniform (BDeu) score. HCS is widely used for its simplicity and computational efficiency, especially when the search space is not prohibitively large [Tsamardinos *et al.* 2006].

Algorithm 1 : Hill Climb Search with BIC Score for BN Structure Learning

Require: Initial BN structure G_0, observed data $\mathcal{D}$
Ensure: Optimal BN structure G

$G \leftarrow G_0$
while stopping criterion is not met **do**
  Generate a set of neighboring BN structures $\{G'_1, G'_2, \ldots, G'_k\}$ by adding, deleting, or reversing an edge in the current structure G
  for each neighboring BN structure G'_i **do**
    Calculate the BIC score $\text{BIC}(G'_i : \mathcal{D})$
  end for
  $G^* \leftarrow \arg\max_{G'_i} \text{BIC}(G'_i : \mathcal{D})$
  if $\text{BIC}(G^* : \mathcal{D}) > \text{BIC}(G : \mathcal{D})$ **then**
    $G \leftarrow G^*$
  else
    break
  end if
end while
return G =0

The Bayesian Information Criterion (BIC) is the scoring metric used to evaluate the network's goodness of fit, while penalising for model complexity. The BIC score is given by:

$$\text{BIC} = -2 \log L + k \log N, \tag{3.2}$$

where L is the model's likelihood, k is the number of parameters, and N is the number of data points.

Algorithm 2 : Simulated Annealing with BIC Score for BN Structure Learning

Require: Initial BN structure G_0, observed data $\mathcal{D}$, initial temperature T_0, cooling rate α
Ensure: Optimal BN structure G

```
G ← G_0
T ← T_0
while stopping criterion is not met do
    Generate a neighboring BN structure G′ by adding, deleting, or reversing an edge in the current
    structure G
    Calculate the change in BIC score ΔBIC = BIC(G′ : D) − BIC(G : D)
    if ΔBIC > 0 then
        G ← G′
    else
        Generate a random number r uniformly from [0, 1]
        if r < exp(ΔBIC/T) then
            G ← G′
        end if
    end if
    Update the temperature: T ← αT
end while
return G =0
```

Simulated Annealing (SA), is a metaheuristic algorithm that probabilistically accepts worse-scoring networks at higher "temperatures" to avoid local optima gradually becoming greedier as it "cools" [De Campos 2011]. SA is particularly useful for navigating complex search landscapes with many local maxima, which are common in high-dimensional BN structure learning tasks involving both meteorological and PV system configuration variables.

Together, these two search algorithms allow for a comparative analysis. HCS offers speed and simplicity, while SA provides robustness against local maxima, a particularly valuable trait when exploring the intricate interdependencies within PV systems. Their score-based nature also allows for the incorporation of domain-specific scoring functions or prior knowledge, which is beneficial for ensuring realistic relationships within the final BN. Several alternatives exist, such as constraint-based algorithms (e.g., the PC algorithm) and hybrid methods (e.g., Max-Min Hill Climbing), each with their strengths. Constraint-based approaches often rely heavily on conditional independence tests, which can become unreliable in high-dimensional or noisy data, typical in environmental modeling [Spirtes *et al.* 2000]. Hybrid methods, while potentially offering better performance, often require more computational resources and complex hyperparameter tuning, making them less suitable for rapid experimentation in the early stages of model development. In contrast, HCS and SA provide a balance between efficiency and model quality.

Optimisations such as Tabu Lists and Random Restarts are incorporated to improve the performance of the Hill Climb Search Algorithm. Tabu Lists prevent the search from revisiting recently explored solutions, thus avoiding cycling and encouraging exploration of new areas in the solution space. Random Restarts involve restarting the search from multiple initial random structures to avoid getting stuck in local optima. Figure 3.5 illustrates a potential issue of utilising the Hill Climb Search algorithm without Random Restarts. Assuming the search procedure starts at the origin of the plane, taking small steps would result in a local optimum. Random Restarts assists the search algorithm to take random steps in the search space, once a local optimum is reached. This assists the algorithm to converge to the global optimum depicted in Figure 3.5.

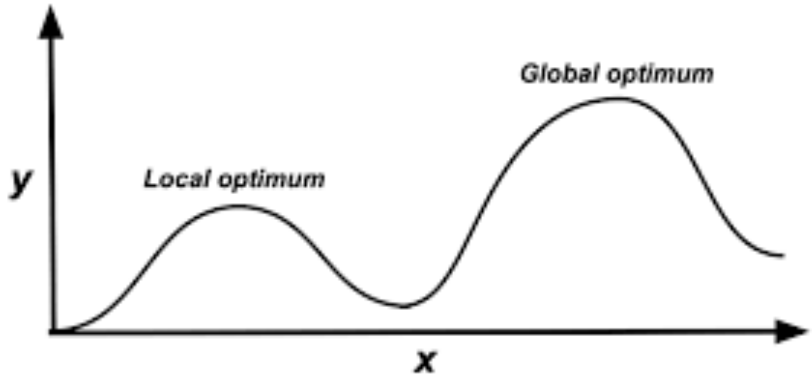

Figure 3.5: The Standard Hill Climb Search May Use Tiny Steps to Unfortunately Converge to a Local Optimum.

Table 3.7: Parameters Optimised in Hill Climb Search

Parameter	Description
`tabu_size`	Size of the tabu list
`max_restarts`	Maximum number of random restarts
`max_iterations`	Maximum iterations per restart

Table 3.8: Parameters Optimised in Simulated Annealing

Parameter	Description
`n_iterations`	Number of iterations for the algorithm
`step_size`	Size of the step in the search space
`initial_temp`	Initial temperature for annealing
`temp_reduction`	Rate of temperature reduction

Learning the Conditional Probability Distributions

Let $\mathcal{D}$ be the observed dataset consisting of N observations. A Bayesian Network is defined by a structure G and a set of Conditional Probability Distributions (CPDs) Θ. Each variable X_j in the network has a CPD $P(X_j \mid \text{Pa}(X_j), \theta_j)$, where $\text{Pa}(X_j)$ denotes the parents of X_j in the network, and θ_j are the parameters of the CPD.

Maximum Likelihood Estimation

MLE aims to determine the set of parameters Θ that maximises the likelihood of the observed data $\mathcal{D}$ given the network structure G. This is analogous to maximising the log-likelihood function:

$$\mathcal{L}(\Theta \mid \mathcal{D}, G) = \log P(\mathcal{D} \mid G, \Theta) = \sum_{i=1}^{N} \sum_{j=1}^{m} \log P(x_j^{(i)} \mid \text{Pa}(x_j^{(i)}), \theta_j). \tag{3.3}$$

To find the MLE estimates of the parameters Θ, we solve the following optimisation problem:

$$\hat{\Theta} = \arg\max_{\Theta} \mathcal{L}(\Theta \mid \mathcal{D}, G). \tag{3.4}$$

Bayesian Estimation with Priors

Bayesian Estimation incorporates prior beliefs about the parameters Θ by using prior distributions. Let $P(\Theta \mid G)$ denote the prior distribution over the parameters Θ, given the network structure G.

The posterior distribution of the parameters Θ, given the data $\mathcal{D}$ and the network structure G, is given by the likelihood of the data given the network structure G and parameters Θ:

$$P(\mathcal{D} \mid G, \Theta) = \prod_{i=1}^{N} \prod_{j=1}^{m} P(x_j^{(i)} \mid \text{Pa}(x_j^{(i)}), \theta_j). \tag{3.5}$$

To perform Bayesian Estimation, we aim to find the posterior distribution $P(\Theta \mid \mathcal{D}, G)$. Often, we summarise this distribution by its mode, known as the Maximum A Posteriori (MAP) estimate:

$$\hat{\Theta}_{MAP} = \arg\max_{\Theta} P(\Theta \mid \mathcal{D}, G). \tag{3.6}$$

Alternatively, the mean of the posterior distribution can be used:

$$\hat{\Theta}_{mean} = \int \Theta P(\Theta \mid \mathcal{D}, G)\, d\Theta. \tag{3.7}$$

By applying these techniques to estimate the CPDs, the methodology constructs a Bayesian Network that accurately represents the probabilistic relationships between PV system variables and environmental conditions in order to maximise PV system output.

MAP estimation is used to determine the optimal PV system parameters by taking into account both the likelihood of the observed data and prior knowledge of the parameters. Given a Bayesian Network that represents the relationships between meteorological variables and the PV system configurations, MAP estimation seeks to provide a single point estimate that maximises the posterior distribution, as shown in Equation 3.1.

3.4.3 Phase 3: Model Evaluation Metrics and Techniques

Metrics

Precision is defined as the ratio of correctly predicted positive observations to all expected positives. It is defined as:

$$\text{Precision} = \frac{TP}{TP + FP},$$

where TP represents the count of true positives and FP denotes the count of false positives.

Recall is the proportion of correctly predicted positive observations to all observations in the actual class. It is defined as:

$$\text{Recall} = \frac{TP}{TP + FN},$$

where FN denotes the count of false negatives.

The F1 metric, calculated as the harmonic mean of Precision and Recall, serves to balance both metrics. It is especially beneficial in scenarios with imbalanced class distributions. The formula is given by:

$$\text{F1} = 2 \cdot \left(\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}\right).$$

Methods

The evaluation of the MAP estimates involve employing forward sampling of the BN to generate approximate samples around these exact estimates. Forward sampling provides a tractable and efficient way to explore the posterior distribution in settings where exact inference is computationally expensive or unnecessary. Given the learned BN structure and parameters, forward sampling enables marginal and joint distributions to be empirically approximated, particularly in high-dimensional models. This is especially useful in PV system optimisation where the model includes both discrete and continuous variables. The corresponding PV system output for the MAP estimates and the samples are plotted to inspect whether the MAP estimates indeed represents the most optimal configuration that results in the maximum PV output, for given meteorological conditions. The optimal Bayesian network's accuracy and prediction performance are evaluated using K-Fold Cross Validation. This approach ensures that the assessment is not influenced by a specific training-validation split and provides a more reliable assessment of the Bayesian Network's capability to generalise unseen data. A paired sample t-test is conducted to measure the statistical significance of employing the optimal configurations per season compared to keeping a constant PV configuration throughout the year.

3.5 Limitations

Bayesian Networks, while powerful, require substantial computational resources to process complex models and large datasets. This can pose challenges for real-time applications or systems operating under constrained computational environments. To manage this limitation and reduce computational complexity, the study limited the collection of data from three key cities in South Africa: Johannesburg, Cape Town, and Durban. By narrowing the geographical scope, the study strikes a balance between computational feasibility and the generalisability of the findings.

Another critical limitation is the potential impact of poor data quality on the model's performance. Inaccurate or incomplete meteorological data can lead to suboptimal forecasts of photovoltaic (PV) output, undermining the decision-making process and system efficiency. To address this issue, the study utilised publicly available meteorological data sourced from PVGIS. This platform is known for its reliability and reputation in providing high-quality, comprehensive meteorological datasets, which are essential for accurate modeling and optimisation.

Finally, the sensitivity of Bayesian Networks to prior assumptions represents a notable challenge. If the priors do not adequately capture real-world conditions, the model may produce biased estimates, leading to inefficient PV system configurations and reduced effectiveness. To mitigate this risk, K-Fold Cross-Validation is employed to rigorously evaluate the model's performance across different subsets of the data. This method provides insights into the robustness of the model and its sensitivity to specific assumptions or data points, ensuring that the conclusions drawn are both reliable and generalisable. These limitations highlight the trade-offs inherent in employing Bayesian Networks for PV system optimisation while underscoring the importance of strategic approaches to mitigate their impact.

3.6 Technical Contributions

This study introduces a novel framework that integrates multiple sources of uncertainty and dependencies in photovoltaic (PV) system performance. Unlike traditional methods, this approach employed Bayesian Networks to model complex interdependencies among variables, leading to more accurate and robust optimisation outcomes. By addressing uncertainties inherent in meteorological conditions and system configurations, the methodology enhanced the reliability of PV output predictions and system designs.

Additionally, the study significantly contributed to improving the adaptability and efficiency of PV systems. The optimisation framework ensured that system configurations are tailored to a wide range of meteorological conditions, enabling PV systems to perform effectively under diverse environmental scenarios. This adaptability translates to enhanced overall performance and long-term reliability, addressing one of the critical challenges in PV system management.

Finally, the research employed a rigorous validation process to evaluate the model's accuracy and robustness. By implementing K-Fold Cross-Validation and other statistical methods, the study ensured that the optimised configurations are both statistically sound and practically viable. This comprehensive validation not only builds confidence in the model's predictions but also sets a new benchmark for modern PV system optimisation techniques, paving the way for further advancements in the field.

Chapter 4

Learning the Structure of the Bayesian Network

Within the discipline of Bayesian structure learning, a variety of techniques can be employed to determine the most effective configurations. This chapter delves into the exploration of two structure learning methods, the Hill Climb and Simulated Annealing search algorithms. These techniques are meticulously applied to discern the most optimal structure, ensuring the highest levels of accuracy and efficiency in the model. Through this comparative analysis, the chapter aims to elucidate the strengths and limitations of each approach, providing valuable insights into their application and performance.

Tables 4.1 and 4.2 illustrate the various parameters that are used to train the Hill Climb Search and Simulated Annealing algorithms.

Tabu List Size	**Random Restarts**	**Search Iterations per Restart**
10	2	10
10	2	20
10	2	50

Table 4.1: Optimised Hill Climb Search Parameters

Iterations	**Step Size**	**Initial Temperature**	**Temperature Reduction**
20	1	1	0.99
40	1	1	0.99
100	1	1	0.99

Table 4.2: Optimised Simulated Annealing Parameters

Based on the analysis and evaluation of the results, the configuration of the Hill Climb Search algorithm with 50 iterations, 2 restarts, and a tabu list size of 10, was found to be the most optimal network. Similarly, for the Simulated Annealing algorithm, the configuration with 100 iterations, a step size of 1, an initial temperature of 1, and a temperature reduction factor of 0.99 was identified as the most effective setup. This selection process ensured that the optimal structure accurately captured the flow of influence amongst variables.

4.1 Structure Learning

4.1.1 Hill Climb Training Analysis

The enhanced Hill Climb Search algorithm (HCS) produced a detailed and robust Bayesian Network structure that effectively captures the probabilistic dependencies among both meteorological variables and PV system configuration parameters. Through iterative local optimisations, the algorithm converged to a network structure that reflects meaningful and interpretable relationships, offering valuable insights into how environmental and system design factors jointly influence PV output. Notably, Figure 4.1 exhibits intuitive and physically consistent dependencies. This level of explainability is particularly beneficial within the energy industry, where domain knowledge must be integrated with data-driven insights for actionable decision-making. The HCS algorithm's ability to uncover such interpretable patterns underscores its utility in structure learning for real-world applications where transparency and trust in model behavior are critical.

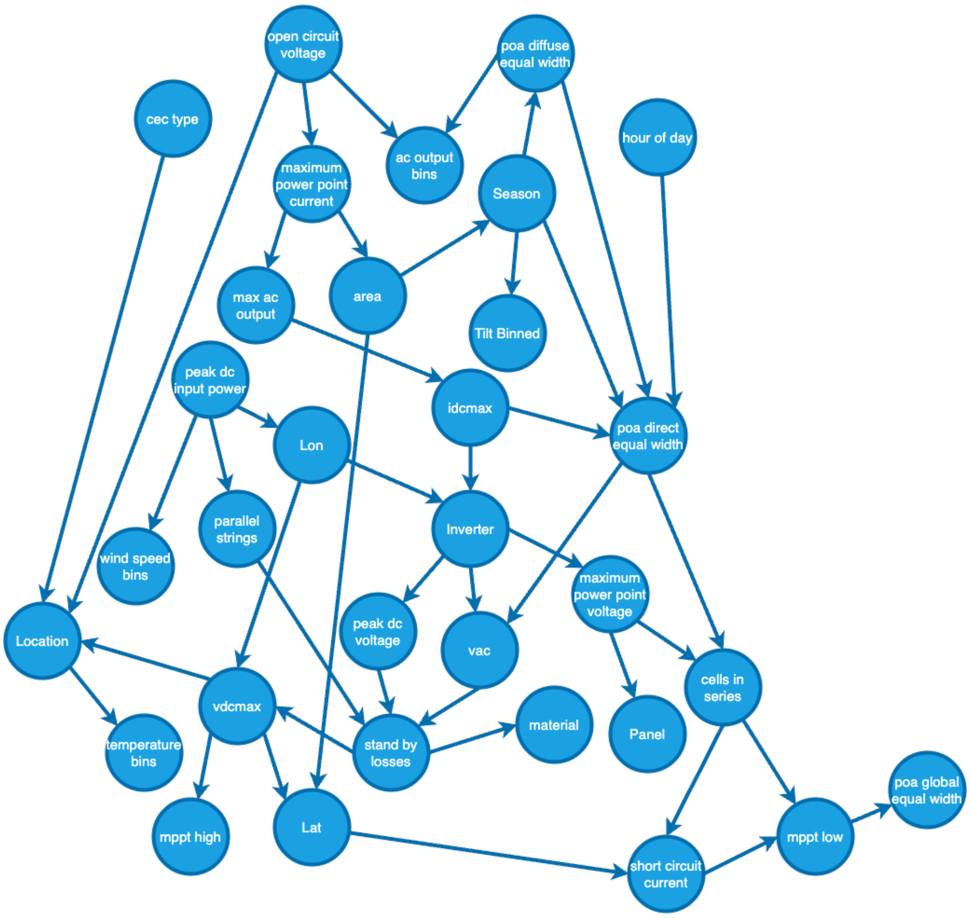

Figure 4.1: Hill Climb Search Bayesian Network Structure

BIC Scores and Computational Time

The Bayesian Information Criterion (BIC) scores illustrated in Figure 4.2 exhibited a clear decreasing trend over the course of 50 iterations. Starting at approximately -25,260,756, the scores progressively decreased to about -8,231,024. The trend is characterised by significant drops in BIC scores during the early iterations, particularly between iterations 1 to 5 and 8 to 10, indicating an initial increase in model complexity. This rapid reduction in BIC scores suggests that the model is quickly improving its fit during the early stages. From iteration 26 onwards, the decrease in BIC scores becomes less dramatic, which indicates that the algorithm is entering a refinement phase where further improvements are more gradual. The time per iteration, illustrated in Figure 4.3 ranged from 69 to 102 seconds. While the computation time remained relatively stable in the early to mid-stages, there was a noticeable increase towards the end, with iterations 30 to 50 taking progressively more time, peaking at approximately 102 seconds. The increase suggested that the algorithm's complexity grows as it approaches the optimal configuration. Restart 2 depicted an increasing trend in the BIC scores, illustrated in Figure 4.2, starting at -19,715,373 and reaching approximately -8,059,082 by the end of the 50 iterations. Notably, significant reductions in BIC scores occurred during the first 10 iterations, reflecting a rapid increase in model complexity, similar to Restart 1. However, the improvements become less pronounced after iteration 19, indicating a more gradual refinement phase as the algorithm optimised the model further. The time per iteration for Restart 2 illustrated in Figure 4.3 ranges from 68 to 97 seconds. As with Restart 1, the early iterations show stable computation times, but there is a slight increase in time towards the later stages. This indicated that, similar to Restart 1, the model required more computational resources as it approached an optimal solution.

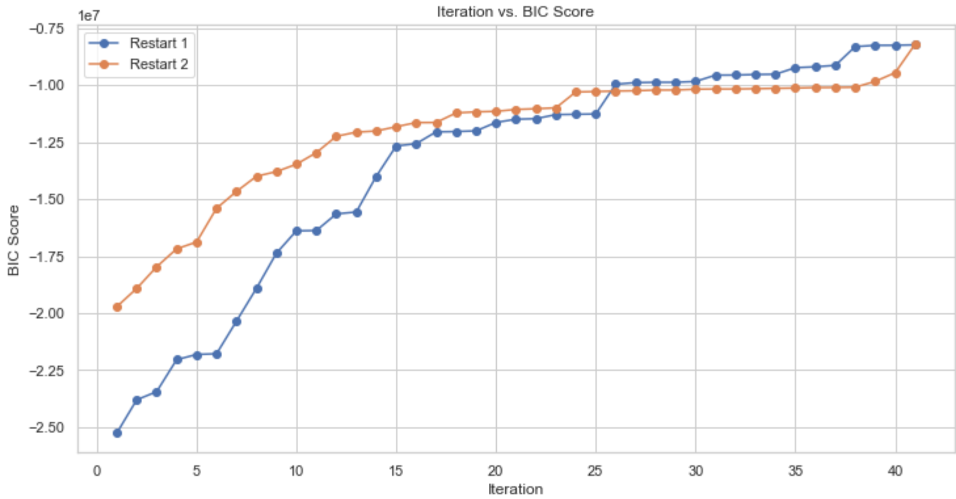

Figure 4.2: The Impact of 50 Iterations on the BIC Score for the Hill Climb Search Algorithm

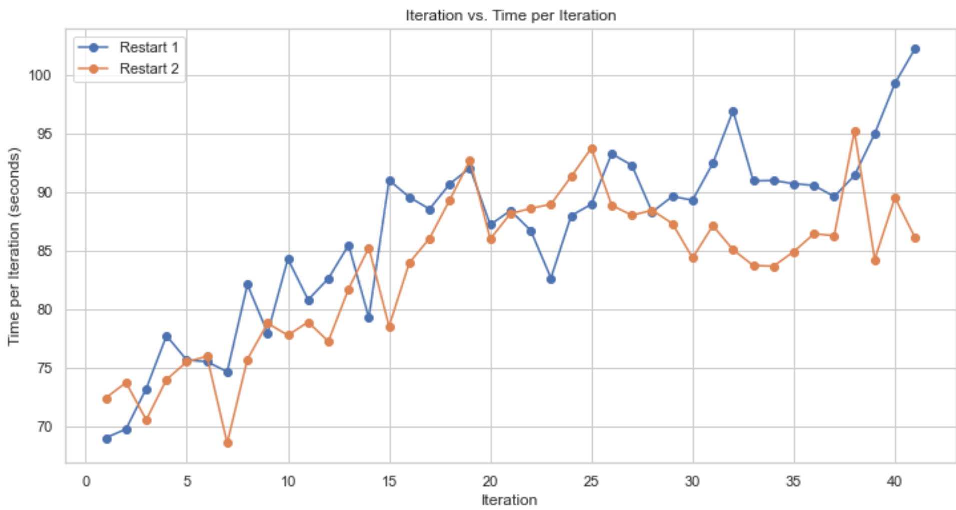

Figure 4.3: The Impact of 50 Iterations on Training Time for the Hill Climb Search Algorithm

Comparative Analysis of Restarts

When comparing the two restarts of the Hill Climb Search, both deployments demonstrated significant increases in BIC scores. However, Restart 2 started with a higher initial BIC score but converged to a similar final value as Restart 1. This suggested that while Restart 2 starts at a higher point, it ultimately achieves comparable results in terms of model performance. Restart 1 generally exhibited longer computation times per iteration compared to Restart 2. Despite this, both restarts showed an increase in time per iteration towards the end of the process, highlighting a common trend of more complex computations as the algorithm refines the model towards an optimal solution.

Insights and Conclusions

Both restarts effectively increase the BIC scores, with significant improvements occurring in the early iterations. This suggests that the early stages of the algorithm are critical for capturing model complexity and rapidly improving the fit. The trade-off between model fit and complexity becomes more evident as the algorithm progresses. Restart 2 demonstrates relatively more efficient performance in terms of computation time while achieving similar increases in BIC scores as Restart 1. This suggests that Restart 2 may be a preferable option in scenarios where computational efficiency is a priority. The slight increase in time per iteration in the later stages of both restarts indicates a higher computational complexity as the algorithm nears an optimal solution. This highlights the increasing computational demands during the refinement phase. These insights suggest that the introduction of Random Restarts offer a slight advantage in computational efficiency, making it a preferable choice in scenarios where computation time is critical. Both restart configurations show the algorithm's robustness and capability to iteratively optimise the BIC scores.

4.1.2 Simulated Annealing Training Analysis

In contrast to the Hill Climb Search, the Simulated Annealing algorithm depicted in Figure 4.4 resulted in a comparatively sparser network structure. While SA is known for its ability to escape local optima through probabilistic acceptance of worse-scoring solutions during the early phases of search, the final structure obtained in this case displayed several limitations. Specifically, certain conditional dependencies that are expected based on domain knowledge were either absent or weakly represented in the network.

This sparsity may be attributed to the algorithm's stochastic nature and sensitivity to the choice of initial temperature, cooling schedule, and scoring function. These hyperparameters critically affect the balance between exploration and exploitation, and suboptimal tuning may lead the algorithm to converge to structures that are statistically plausible but semantically or physically unrealistic. As a result, some discovered relationships lacked interpretability or contradicted known causal mechanisms in photovoltaic system behavior.

Furthermore, while sparse structures can be advantageous in reducing model complexity and enhancing generalisation, excessive sparsity compromises the network's ability to capture meaningful joint distributions, thereby limiting its effectiveness for both inference and decision support. In this context, the SA-derived network was less effective in supporting reliable predictive reasoning and explaining the underlying dynamics of PV system performance under varying meteorological conditions. These findings underscore the importance of algorithm selection and parameter calibration in structure learning, particularly when the model is expected to serve both predictive and explanatory roles in critical applications such as PV system optimisation.

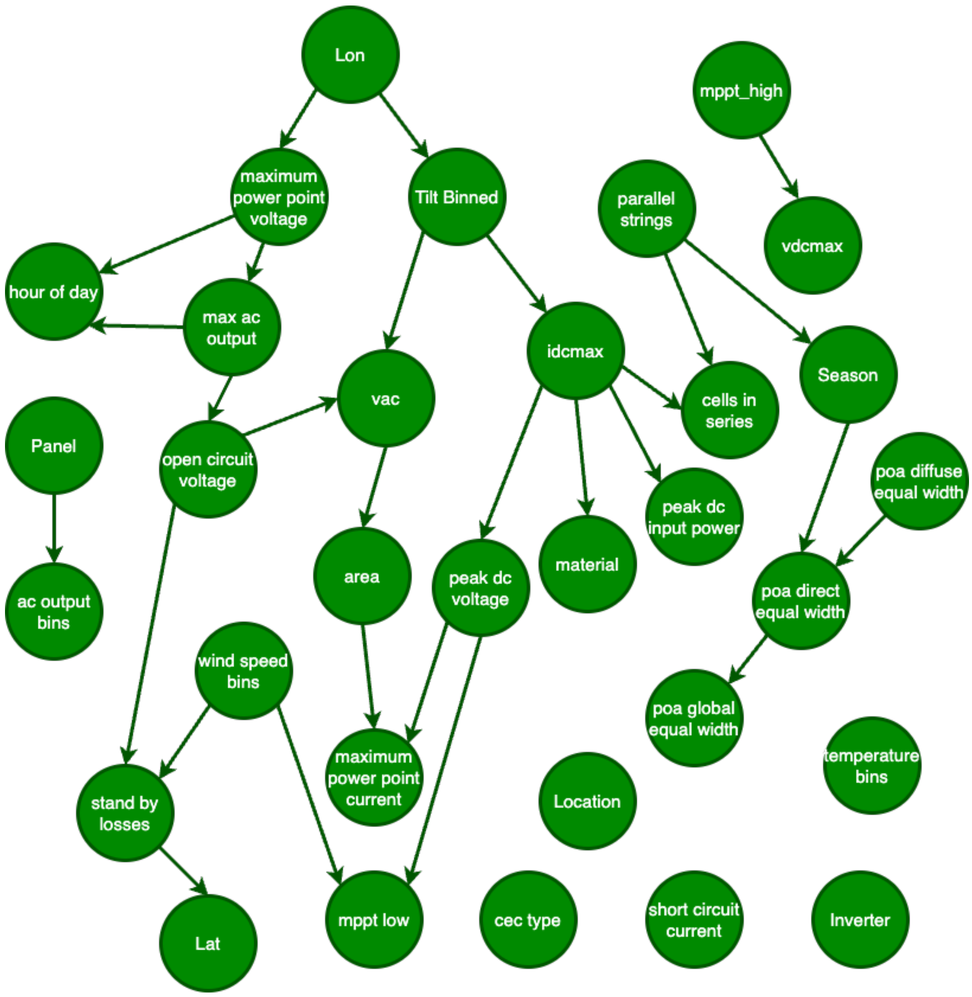

Figure 4.4: Simulated Annealing Bayesian Network Structure

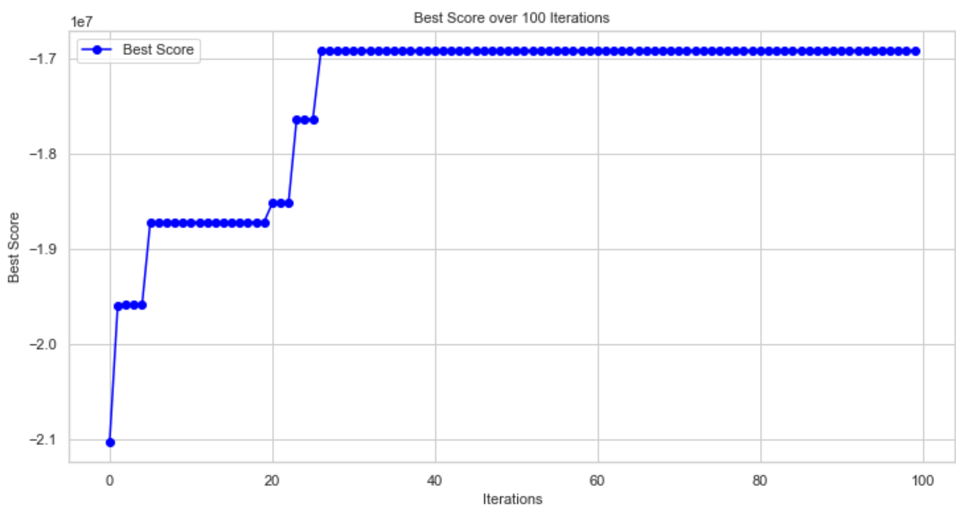

Figure 4.5: The Impact of 100 Iterations on the BIC Score for the Simulated Annealing Search Algorithm

BIC Scores and Computational Time

The BIC scores illustrated in Figure 4.5 showed a clear increasing trend over 100 iterations. Starting from an initial value of approximately -21,037,944, the scores steadily increase and converge to around -16,915,049. The most significant increases occur during the initial phase, particularly between iterations 0 and 5. This displayed rapid optimisation of model complexity at the beginning of the process. After iteration 5, the improvements in BIC scores became more gradual, reflecting a transition to a refinement phase where the algorithm fine-tunes the model with smaller, incremental gains. By iteration 26, the score stabilised near -16,915,049, with only minor fluctuations observed thereafter, indicating that the algorithm has reached convergence. The computation time per iteration depicted in Figure 4.6 for the SA algorithm ranged between 5.89 seconds and 7.40 seconds. Throughout the 100 iterations, the time remained relatively stable, with only minor fluctuations. A notable exception occurs at iteration 95, where the computation time spiked to 7.40 seconds, possibly due to increased complexity or additional processing requirements at this specific step. Despite this anomaly, the overall trend suggested consistent computational demands across the iterations.

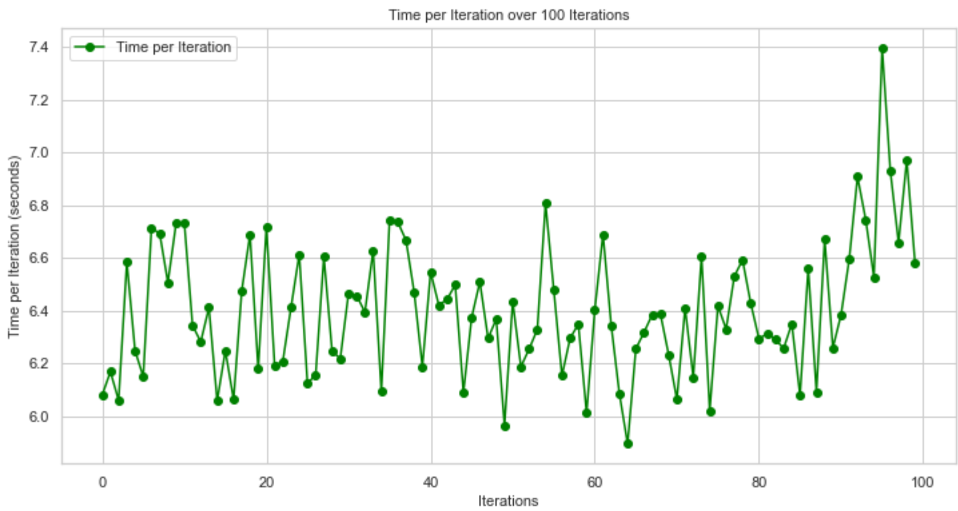

Figure 4.6: The Impact of 100 Iterations on Training Time for the Simulated Annealing Search Algorithm

Insights and Conclusions

The algorithm demonstrated consistent computational stability, with only minor fluctuations in time per iteration. This stability makes it a reliable choice for tasks requiring predictable computational performance. The significant increases in BIC scores during the initial iterations underscored the algorithm's capability to rapidly optimise the model, balancing fit and complexity effectively. The plateau observed from iteration 26 onwards confirmed that the algorithm converges to a near-optimal solution within a reasonable number of iterations. These insights indicate that the Simulated Annealing algorithm is effective in achieving rapid optimisation and demonstrates stability in both performance and computational requirements over extended iterations.

4.2 Performance and Accuracy

Hill Climb Cross Validation Results

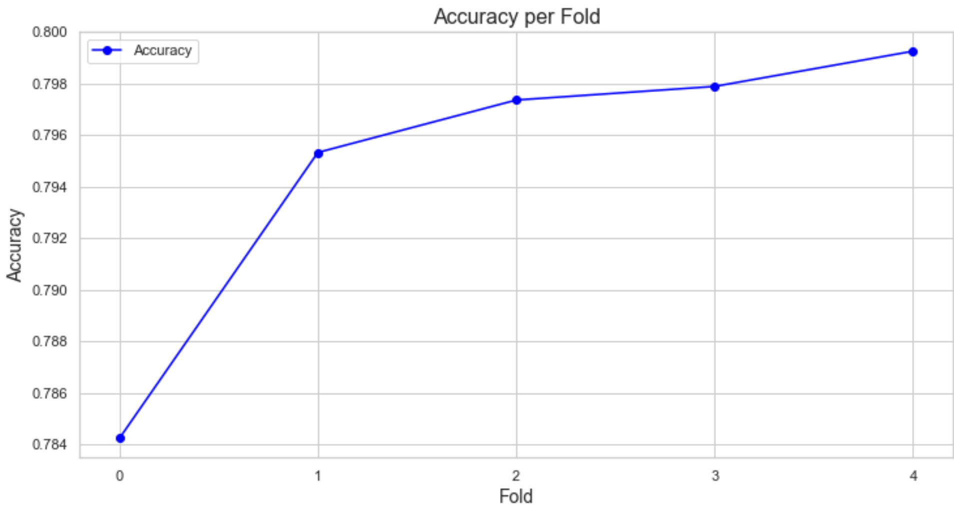

Figure 4.7: Hill Climb Search Cross Validation Accuracy

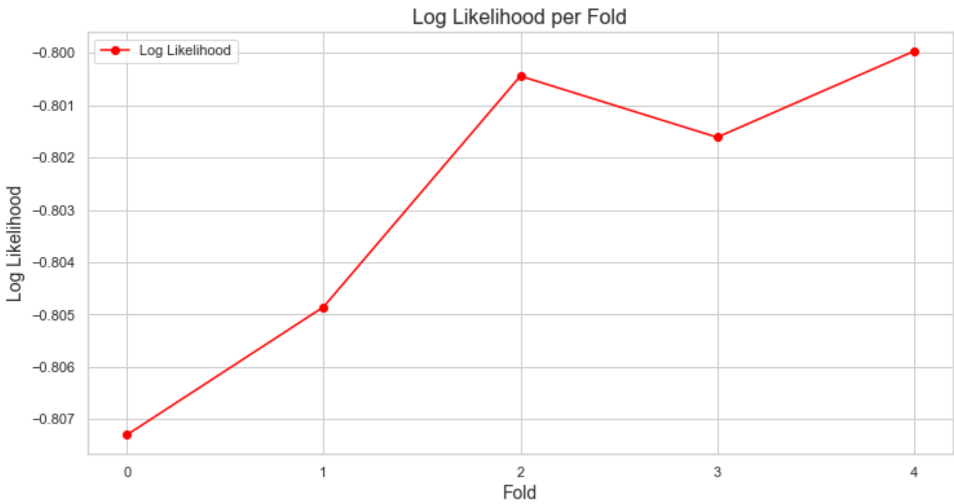

Figure 4.8: Hill Climb Search Cross Validation Log Likelihood

The analysis of the Hill Climb Search cross-validation results depicted in Figures 4.7 and 4.8 showed a consistent improvement in both accuracy and log-likelihood metrics across the five validation folds. The accuracy steadily increased from 0.7842 in fold 0 to 0.7993 in fold 4, indicating a positive trend in model performance. Similarly, the log likelihood improved from -0.8073 in fold 0 to -0.8000 in fold 4, suggesting a better fit of the model to the data. These improvements highlight the model's reliability and robustness in generalising to different data subsets. The highest accuracy and best log likelihood values observed in fold 4 suggest that this fold provided the most representative data for training. Overall, the model demonstrated effective learning and consistent performance enhancement, indicating a well-trained model with good generalisation capabilities.

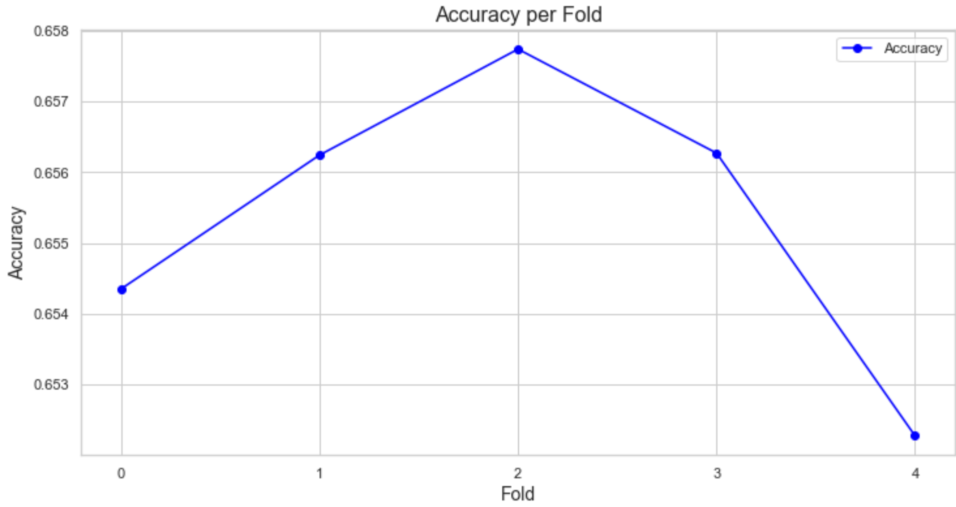

Figure 4.9: Simulated Annealing Cross Validation Accuracy

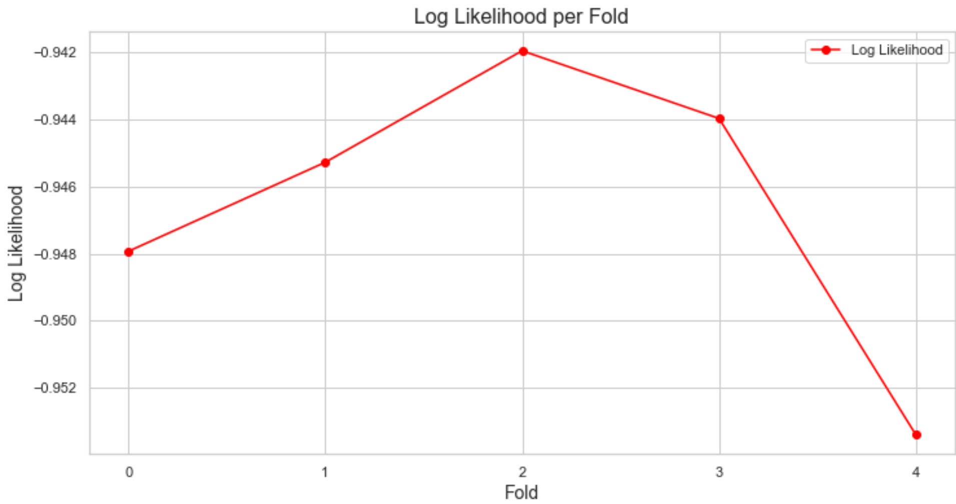

Figure 4.10: Simulated Annealing Cross Validation Log Likelihood

Simulated Annealing Cross Validation Results

The analysis of the Simulated Annealing results illustrated in Figures 4.9 and 4.10 revealed a consistent performance across the five validation folds, with minor fluctuations in accuracy ranging from approximately 0.652 to 0.658 and log likelihood values from around -0.954 to -0.942. In comparison, the Hill Climb Search results showed a slightly higher and more consistent performance with accuracy ranging from approximately 0.784 to 0.799 and log likelihood values from around -0.820 to -0.725. The Hill Climb Search algorithm demonstrated a more pronounced improvement in both accuracy and log likelihood per fold, suggesting a better overall fit and performance. However, both algorithms showed reliable generalisation capabilities and stable results across folds, with Hill Climb Search having a slight edge in terms of higher accuracy and better log likelihood values.

Metrics

Metric	Hill Climb Search	Simulated Annealing
Accuracy	0.82	0.65
Precision	0.83	0.58
Recall	0.79	0.57
F1 Score	0.81	0.575

Table 4.3: Comparison of Metrics for Hill Climb Search and Simulated Annealing

The analysis of both Hill Climb Search and Simulated Annealing results highlights significant differences in their performance metrics, evident in Table 4.3. The Hill Climb Search showed higher and more consistent accuracy (0.82) compared to Simulated Annealing (0.655), along with better precision (0.83) and recall (0.79). This resulted in a higher F1 score of approximately 0.81 for Hill Climb Search versus 0.575 for Simulated Annealing, indicating superior overall performance and reliability in Hill Climb Search. On the other hand, Simulated Annealing presented more variability but still maintained a reasonable level of accuracy and precision. The steadier performance of Hill Climb Search in terms of accuracy, precision, recall, and log likelihood suggested it is more robust and effective for the given dataset. In contrast, while Simulated Annealing showed potential, its slightly lower metrics indicated it may require further tuning to achieve optimal performance.

4.3 The Optimal Network Structure

In comparing Hill Climb Search with Tabu Lists and Random Restarts to Simulated Annealing (SA), it is evident that the Hill Climb Search demonstrated superior performance and reliability. Hill Climb Search consistently achieved superior performance and accuracy when compared to Simulated Annealing. The inclusion of Tabu Lists in Hill Climb Search prevented revisiting previous solutions, enhancing efficiency and avoiding local optima, while Random Restarts allowed for broader exploration of the solution space. These features make Hill Climb Search more effective at identifying globally optimal solutions. The consistency in the Hill Climb Search's performance metrics across different validation folds underscored its stability and robustness which are crucial for real-world applications. In contrast, Simulated Annealing showed more variability and slightly lower metrics, suggesting it might not always explore the solution space as effectively. Overall, the Hill Climb Search with Tabu Lists and Random Restarts resulted in the most optimal network structure due to its superior accuracy, robustness, and reliability in generalisingnew data.

Chapter 5

MAP Estimation to Maximise PV System Output

5.1 Maximum A Posteriori (MAP) Query

In this chapter, using the optimal Hill Climb Bayesian Network, we conduct Maximum A Posteriori (MAP) Estimation to solve for the most optimal seasonal configuration of the network variables, given specific evidence. This process involves assessing the performance of the algorithm under different conditions, including various seasons (Summer, Autumn, Winter, Spring) and geographical locations (Cape Town, Johannesburg, Durban).

The generalised MAP query for each location and season is:

$$\hat{\mathbf{X}} = \arg\max_{\mathbf{X}} P(\mathbf{X}|\{\text{ac\_output\_bins} = \text{Peak Output}, \text{season}, \text{Location}\}), \tag{5.1}$$

where season $\in$ $\{\text{Summer}, \text{Autumn}, \text{Winter}, \text{Spring}\}$, Location $\in$ $\{\text{Cape Town}, \text{Johannesburg}, \text{Durban}\}$ and X represents the PV system configuration.

Table 5.1: Seasonal MAP Estimates for Johannesburg, Cape Town, and Durban

Location	**Season**	**MPP Voltage (V)**	**Inverter**	**Panel**	**Tilt Binned**
Johannesburg	Summer	40.762	Power One PVI 6000	SunPower SPR 220	100
	Winter	40.762	Power One PVI 6000	SunPower SPR 220	39
	Autumn	40.762	Power One PVI 6000	SunPower SPR 220	63
	Spring	40.762	Power One PVI 6000	SunPower SPR 220	21
Cape Town	Summer	50.000	Growatt 3600 MTL	Schott ASE 300	103
	Winter	50.000	Growatt 3600 MTL	Schott ASE 300	42
	Autumn	50.000	Growatt 3600 MTL	Schott ASE 300	65
	Spring	50.000	Growatt 3600 MTL	Schott ASE 300	18
Durban	Summer	17.600	Power One PVI 3.6	BP Solar SX365	103
	Winter	17.600	Power One PVI 3.6	BP Solar SX365	35
	Autumn	17.600	Power One PVI 3.6	BP Solar SX365	67
	Spring	17.600	Power One PVI 3.6	BP Solar SX365	24

5.2 Analysing MAP Estimates

After sampling 10,000 data points using forward sampling, the MAP estimates consistently outperformed the samples in terms of average seasonal PV system output produced. The results, presented in Tables 5.2 and 5.3, demonstrate the consistent superior performance of the MAP estimates over a substantial number of forward samples, suggesting that the MAP estimates are not only robust but also likely represent the global optimum. The ability of MAP to leverage prior information, integrate observed data, and yield stable, accurate results further reinforces its superiority in achieving the most probable and optimal parameter values.

Table 5.2: MAP Estimates Seasonal Output Mean for Different Locations

Season	Location	Output Mean (Watts)
Autumn	Cape Town	203.918877
	Durban	75.294256
	Johannesburg	203.588477
Spring	Cape Town	501.286595
	Durban	197.494309
	Johannesburg	460.277850
Summer	Cape Town	753.091865
	Durban	336.019858
	Johannesburg	578.311162
Winter	Cape Town	121.265447
	Durban	41.882158
	Johannesburg	90.380451

Table 5.3: Forward Sampling Seasonal Output Mean for Different Locations

Season	Location	Output Mean (Watts)
Autumn	Cape Town	76.918752
	Durban	66.225465
	Johannesburg	23.610496
Spring	Cape Town	217.733716
	Durban	154.726212
	Johannesburg	124.011874
Summer	Cape Town	349.058961
	Durban	227.950413
	Johannesburg	173.988087
Winter	Cape Town	40.087780
	Durban	27.043333
	Johannesburg	20.543517

5.2.1 Seasonal Performance Comparison

The highest PV system output is observed across all locations, with MAP estimates significantly surpassing the forward sampling output values. For instance, Cape Town's Summer MAP estimate (753.09 watts) is more than double its forward sampling counterpart (349.06 watts), underscoring the effectiveness of MAP in capturing the best possible PV system configurations. However, the large discrepancy suggests that forward sampling might not be effectively capturing high-output scenarios, potentially due to under-exploration of optimal configurations. The lowest PV output is recorded during Winter, with both MAP and forward sampling showing substantial reductions. However, the MAP estimates still maintain a notable advantage, such as in Johannesburg, where the MAP estimate (90.38 watts) is more than four times higher than the forward sampling result (20.54 watts). While the gap between MAP and forward sampling results remains substantial, the trend of MAP outperforming forward sampling persists. For instance, in Cape Town during Spring, the MAP estimate (501.29 watts) more than doubles the forward sampling mean (217.73 watts).

The results indicate that MAP estimates significantly outperform forward sampling across all seasons and locations. This superior performance can be attributed to several factors. Firstly, MAP estimation effectively integrates prior knowledge with observed data, enabling the computation of the most probable parameter values and ensuring that the inferred PV system configurations are optimal. Unlike forward sampling, which produces a broad range of possible values due to inherent stochastic variability, MAP estimation provides stable and consistent results. Furthermore, the accuracy of MAP estimates is notably higher, as they capture the underlying patterns of seasonal variations in PV system output more effectively than forward samples, which are often susceptible to randomness and fluctuations. These findings are further visualised in Figure 5.1, which illustrates the seasonal output means obtained through MAP estimates and forward sampling.

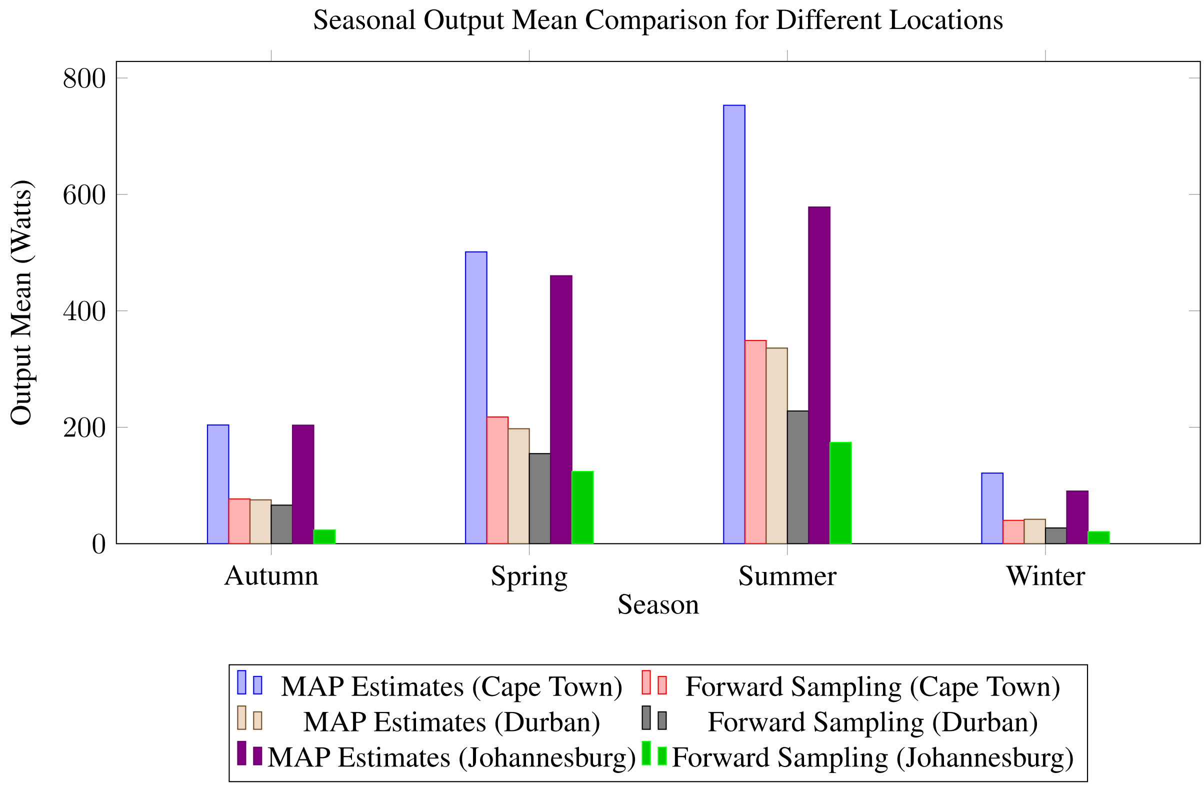

Figure 5.1: MAP Estimates Outperforming Forward Sampling Seasonal Output Mean for Different Locations

The persistent underperformance of forward sampling highlights its limitations in navigating the parameter space efficiently. The stochastic nature of forward sampling leads to suboptimal configurations due to its reliance on random exploration rather than directed optimisation. This inefficiency is particularly evident in scenarios with complex dependencies, such as PV system configurations influenced by dynamic meteorological conditions. MAP estimation, on the other hand, leverages both prior knowledge and observed data to converge to the most probable and optimal system parameter values. The results strongly support the argument that MAP estimation is superior in identifying optimal solutions.

5.3 Conclusion

This analysis demonstrates the effectiveness of MAP estimation in optimising seasonal PV system output. By leveraging prior information and observed data, MAP estimates consistently outperform forward sampling, producing higher and more stable system output means across all locations and seasons. These findings underscore the advantages of Bayesian inference in PV system optimisation, offering a powerful and novel approach for improving PV system energy yield. The robustness of MAP estimation ensures that decision making in renewable energy applications is not only data driven but also more reliable and adaptable to seasonal variations. As the demand for efficient renewable energy solutions grows, the role of Bayesian techniques in enhancing forecasting accuracy and system optimisation becomes increasingly valuable.

Chapter 6

Statistical Significance of MAP Estimates

This experiment compares the photovoltaic (PV) output generated by Maximum A Posteriori (MAP) estimates with that produced by fixed configurations on a seasonal basis. The experiment assists with determining whether MAP estimates provide a statistically significant increase in PV output across different seasons and locations.

6.1 Experimental Design

The following section provides a comprehensive discussion of the key components of the experimental design. We begin by clearly defining the hypothesis, which serves as the foundation for the experiment. The hypothesis outlines the expected relationship between MAP-based configurations and PV output, allowing us to establish a formal statistical framework for analysis. We then detail the independent and dependent variables involved in our experiment. Independent variables, such as configuration type, seasonal variations, and location, play a crucial role in influencing PV output, while the dependent variable, PV output, serves as the primary metric for performance evaluation. Additionally, we present the experimental setup, describing the methodology and approach taken to ensure robust and reliable results. By carefully structuring our study design, the experiment provides meaningful insights into the benefits of Bayesian Network-based optimisation for PV systems.

6.1.1 Declaring the Experimental Hypothesis

The goal of the experiment is evaluated by establishing the following hypotheses:

1) Null Hypothesis (H_0): There is no significant difference in PV output between MAP estimates and fixed configurations. This implies that using a probabilistic approach to determine optimal PV configurations does not yield superior results when compared to traditional fixed setups. The null hypothesis serves as a benchmark to establish whether the proposed method introduces meaningful improvements. A rejection of the hypothesis would demonstrate the effectiveness of MAP estimates in optimising PV system performance.

2) Alternate Hypothesis (H_1): MAP estimates provide significantly higher PV output than fixed configurations. This hypothesis indicates that Bayesian Network-based optimisation of PV configurations lead to tangible performance gains.

6.1.2 Variables

In this experiment we investigate the relationship between various independent variables and PV output, which serves as the dependent variable. The independent variables have been carefully selected to capture the key factors influencing PV system performance, ensuring that the analysis accounts for critical environmental and configuration-based considerations. A detailed discussion of each category of variables included in the experiment is provided below.

Independent Variables

The independent variables in the experiment encompass three primary factors, namely configuration type, seasonal variations, and geographic location. These factors have a direct influence on the performance of a PV system and are critical for optimising energy output.

Configuration Type captures two distinct approaches to PV system configuration. These are systems using Maximum A Posteriori (MAP) estimates and fixed configurations. The MAP estimates are derived using Bayesian inference techniques, incorporating probabilistic modeling to determine the most likely system configurations under given seasonal meteorological conditions. In contrast, fixed configurations rely on predetermined, static settings that do not adapt dynamically to changing environmental variables. By comparing these two approaches, the experiment evaluates the effectiveness of probabilistic optimisation against conventional fixed parameter settings.

The study considers seasonal variations, specifically categorising data into Summer, Autumn, Winter, and Spring. Seasonal changes affect critical meteorological variables such as solar irradiance, temperature, and wind speed, all of which influence PV performance. For instance, Winter months may experience lower irradiance due to shorter daylight hours, whereas Summer months may present higher temperatures, potentially reducing PV efficiency due to thermal losses. By including seasonal effects, the experiment assesses how configuration strategies perform under varying environmental conditions.

For the purposes of simplicity and computational efficiency, the experiment focuses on a single location, Johannesburg. The analytical framework and underlying logic of the study are generalisable and can be extended to other regions, where similar outcomes are anticipated under comparable conditions.

Dependent Variable

The primary dependent variable in this study is PV output, which represents the amount of electrical energy generated by the PV system under different configurations and seasonal conditions. PV output serves as the key performance indicator for evaluating system effectiveness. By analysing the impact of different independent variables on PV output, the study derives insights into the optimal configuration approach which maximises energy generation.

Overall, the structured consideration of these variables ensures a comprehensive assessment of PV system performance, providing a robust framework for evaluating the benefits of adaptive configuration strategies over static alternatives.

6.1.3 Experimental Procedure

To systematically evaluate the performance of different PV configurations, we follow a structured experimental procedure. This approach ensures a robust comparison between MAP-based and fixed configurations under real-world conditions in Johannesburg.

We utilise the MAP query to determine the optimal system configurations for each season and location. This involves leveraging Bayesian inference to identify the most probable settings which maximise PV output given historical seasonal weather patterns and PV system configurations. Both MAP-based and fixed configurations are simulated under real-world conditions using PVlib, a widely used Python library for PV performance modeling. These simulations allow for a controlled comparison, ensuring that differences in PV output are attributable to the configuration method rather than extraneous factors. PV output is measured and recorded for each configuration type over the duration of each season. This data collection process enables an empirical assessment of system performance, facilitating a direct comparison between MAP-optimised and fixed configurations. By systematically implementing these experimental steps, we ensure a rigorous evaluation of PV system configurations, providing insights into the benefits and limitations of adaptive versus static optimisation approaches.

6.1.4 Statistical Analysis

A paired t-test is conducted for each season to compare the mean PV output of MAP estimates and fixed configurations. This test is appropriate as it accounts for the paired nature of the observations, where each configuration type is evaluated under identical seasonal conditions. The p-value is calculated to assess the statistical significance of the observed differences. The p-value provides insight into whether the performance variations between MAP-based and fixed configurations are statistically meaningful. A significance level of $\alpha = 0.05$ is used, ensuring that results are evaluated with a 95% confidence level. If $p \leq \alpha$, we reject the null hypothesis and conclude that MAP estimates provide significantly higher PV system output. However if $p > \alpha$, we fail to reject the null hypothesis and conclude that there is no significant difference in PV system output.

The paired t-test is computed using the following formula:

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}, \tag{6.1}$$

where $\bar{d}$ is the mean of the differences between paired observations, s_d is the standard deviation of the differences, and n is the number of pairs. This test allows us to quantify the effect of configuration strategy on PV output and determine whether MAP-based configurations yield a statistically significant improvement over fixed configurations.

6.2 Analysis of Paired t-Test Results for Johannesburg

The results from the paired t-tests for Johannesburg across different seasons reveal statistically significant differences between the dynamic and fixed PV system output configurations. Each season demonstrates varying degrees of improvement when utilising the dynamic configuration, emphasising its effectiveness in optimising PV system output.

During Summer, the t-statistic of 41.32, accompanied by a p-value of 0.0, signifies an indisputable improvement in PV system output with the dynamic configuration. This substantial difference suggests that in the high-radiation summer months, the adaptability of the dynamic system allows it to harness solar energy far more efficiently than a fixed system, leading to considerably higher energy yields.

Similarly, in Autumn, the t-statistic remains exceptionally high at 35.05, with a p-value of approximately 4.97e-247, effectively zero. This result reinforces the conclusion that the dynamic configuration consistently outperforms the fixed alternative, even as solar angles and weather conditions shift. The continued superiority of the dynamic system in autumn highlights its ability to adjust effectively to seasonal variations in sunlight exposure.

Winter results also demonstrate a significant difference, though with a slightly lower magnitude compared to Summer and Autumn. The t-statistic of 20.47 and an extremely low p-value of 2.38e-90 indicate that while the dynamic configuration still yields a higher PV system output, the relative improvement is less pronounced. This could be attributed to shorter daylight hours and a lower solar elevation angle, where the advantage of adjustability, while still beneficial, is somewhat diminished.

Spring, however, exhibits the highest statistical differentiation among all seasons, with a t-statistic of 45.72 and a p-value of 0.0. This suggests that the dynamic configuration provides the most significant gains in PV system output during this season. The sharp increase in solar availability and more favorable angles likely contribute to the maximisation of energy captured, further validating the effectiveness of the dynamic approach.

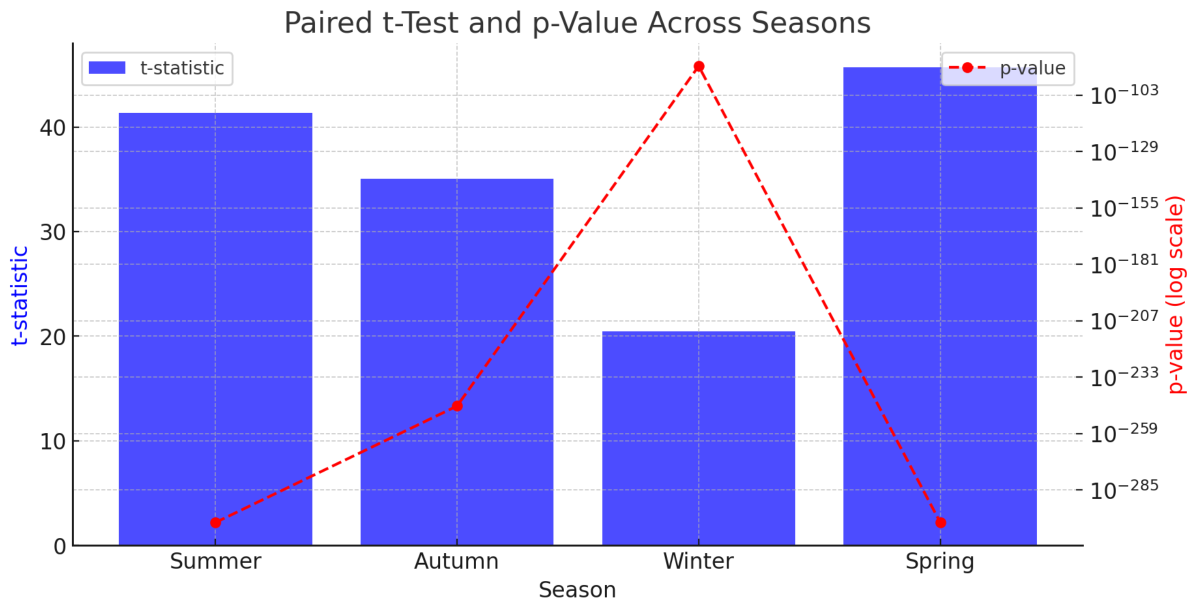

Figure 6.1: t-Test and p-Values across Seasons

6.3 Conclusion

Overall, the statistical results across all seasons underscore the substantial benefits of using dynamic PV configurations over fixed ones. The consistently high t-statistics and near-zero p-values confirm that the observed improvements are not due to random variations but rather to the superior adaptability of the dynamic system. These findings strongly advocate for the use of dynamic tilt configurations to maximise PV efficiency, particularly in regions with varying seasonal solar conditions such as Johannesburg.

Chapter 7

Conclusion and Future Work

The objective of this study is to maximise PV system output by using Bayesian Networks to simulate and analyse the interaction of meteorological variables and PV system configurations. The study's aims are centered around establishing optimal network structures with structure learning algorithms, using MAP Estimation to select the most optimal seasonal PV system configurations, and analysing the statistical significance of dynamic configurations versus fixed configurations.

7.1 Addressing the Research Questions

The study effectively and accurately examined and answered the following research questions:

Which structure learning procedure optimises the likelihood of PV system output given the dependencies between the meteorological and PV system variables?

Using structure learning techniques, the study demonstrated that the Hill Climb Search Algorithm using a Tabu List with Random Restarts outperformed Simulated Annealing in terms of producing network structures that successfully captured variable dependencies. The resilience and effectiveness of Hill Climb Search make it a dependable solution for similar applications.

The BIC score of the Hill Climb Search converged to -8,059,082, whereas Simulated Annealing's BIC equated to -16,915,049. Although Simulated Annealing exhibits a higher BIC score, it results in a sparse network with some nodes excluded from the final structure. In contrast, Hill Climb Search yields a more intricate and accurate graph, effectively capturing the complex relationships among variables. This complexity is vital for creating models that accurately reflect the underlying data structure. The capability of Hill Climb Search to produce a detailed and accurate network is crucial for real-world applications, where a comprehensive understanding of variable interactions is essential. This detailed representation ensures that models derived from Hill Climb Search are robust and more applicable to practical scenarios.

- Accuracy: Hill Climb Search achieves an accuracy of 82%, significantly higher than Simulated Annealing's 65%. This considerable difference underscores Hill Climb Search's enhanced ability to correctly classify instances, thereby offering a more reliable model for prediction tasks.
- Precision: Hill Climb Search demonstrates a precision of 83%, compared to 58% for Simulated Annealing. This indicates that Hill Climb Search is more effective in minimising false positives, thus ensuring a higher proportion of correct positive predictions.

- Recall: With a recall of 79%, Hill Climb Search outperforms Simulated Annealing's 57%, highlighting its superior capability in identifying true positives and reducing the number of false negatives.
- F1 Score: The F1 score of 81% for Hill Climb Search, versus 58% for Simulated Annealing, further accentuates Hill Climb Search's balanced performance in terms of precision and recall, making it a more comprehensive measure of the algorithm's efficacy.

Hill Climb Search emerges as the superior approach for constructing network structures, evidenced by its higher accuracy, precision, recall, and F1 score. Despite the lower BIC score of Simulated Annealing, its tendency to produce sparse and potentially incomplete networks undermines its effectiveness. Hill Climb Search, by producing more comprehensive and accurate graphs, better captures the complexity of relationships within the data. Consequently, for applications requiring high accuracy and a nuanced understanding of variable interactions, Hill Climb Search is the preferred algorithm.

What is the most likely seasonal configuration of PV system variables which maximises PV system output?

The implementation of MAP Estimation within the learned Bayesian Network successfully identify PV configurations which maximise seasonal system output. The MAP estimates reliably provide global maxima in the context of the dataset, demonstrating the potential of Bayesian approaches for decision-making under uncertainty.

What is the impact of implementing optimal PV system configurations per season, compared to maintaining a fixed configuration throughout the year?

The statistical experiment demonstrates that dynamic configurations produce greater PV system output compared to fixed configurations. This conclusion emphasises the need for adaptive techniques in PV system optimisation, especially in regions with changing meteorological circumstances.

7.2 Insights and Implications

The study's findings demonstrate the efficacy of Bayesian Networks as a tool for optimising PV system performance. The combination of structure learning, MAP Estimation, and statistical procedures creates a comprehensive framework for making data-driven decisions in renewable energy systems. The improved performance of Hill Climb Search highlights the relevance of structure learning in Bayesian Network applications. Furthermore, the demonstrated advantages of dynamic configurations provide compelling evidence for their use in practical PV system designs.

7.3 Limitations

Despite the promising results, this research is subject to certain limitations which warrant consideration. Firstly, the findings are highly dependent on the quality and scope of the dataset used. Insufficient temporal or spatial coverage in the data may limit the generalisability of the results, potentially restricting their applicability to broader contexts or varying environmental conditions.

Secondly, the computational complexity of the methodology presents a significant challenge. Both the structure learning process and Maximum A Posteriori (MAP) Estimation are computationally intensive. Training the optimal Bayesian Network required a considerable amount of time, primarily influenced by the dataset's size. This intensity may limit scalability to larger datasets or more intricate network structures. Addressing this issue in future studies could involve leveraging machines with greater computational power to reduce runtime and improve efficiency.

Lastly, Certain meteorological and system variables are simplified, which may overlook nuanced interactions present in real-world scenarios. Recognising these limitations provides a foundation for refining and expanding this work, ensuring its continued relevance and applicability in practical settings.

7.4 Future Work

Building on the findings and limitations of this study, several avenues for future research emerge that can enhance the methodology and expand its applicability. One promising direction is the exploration of enhanced structure learning algorithms. Advanced techniques offer potential for improving the accuracy and scalability of Bayesian Network models. These methods could help strike a balance between computational efficiency and model accuracy, making the approach more robust and versatile.

Additionally, incorporating a broader range of variables into the analysis represents a key opportunity. By integrating more granular meteorological data and detailed PV system parameters, future studies could capture more complex interdependencies that influence system performance. This enrichment would likely improve the model's predictive reliability and its alignment with real-world dynamics.

Another area for development is the creation of AI frameworks for real-time dynamic optimisation. Leveraging AI and live meteorological data to reconfigure PV systems in real-time has the potential to maximise power output and enhance system adaptability. Testing and implementing such dynamic frameworks would mark a significant step forward in the practical application of Bayesian Networks.

Finally, addressing scalability and efficiency issues is critical for expanding the methodology's potential. Future research could investigate methods to reduce computational overhead, such as employing parallel computing techniques or developing advanced machine learning pipelines. These innovations could enable the application of Bayesian Networks within larger, more complex systems without compromising performance or runtime. By pursuing these research directions, the field can advance towards more effective, efficient, and adaptable optimisation of photovoltaic systems, ultimately contributing to the broader adoption of renewable energy solutions.

Appendix A

Tilt Bins

The following table presents the discretisation scheme applied to the continuous tilt angle values. The tilt angles, ranging from 1° to 44°, were categorised into 113 discrete bins to facilitate structure learning within the Bayesian Network framework.

Bin Number	Range
0	(0.999, 1.377]
1	(1.377, 1.754]
2	(1.754, 2.132]
3	(2.132, 2.509]
4	(2.509, 2.886]
5	(2.886, 3.263]
6	(3.263, 3.64]
7	(3.64, 4.018]
8	(4.018, 4.395]
9	(4.395, 4.772]
10	(4.772, 5.149]
11	(5.149, 5.526]
12	(5.526, 5.904]
13	(5.904, 6.281]
14	(6.281, 6.658]
15	(6.658, 7.035]
16	(7.035, 7.412]
17	(7.412, 7.789]
18	(7.789, 8.167]
19	(8.167, 8.544]
20	(8.544, 8.921]
21	(8.921, 9.298]
22	(9.298, 9.675]
23	(9.675, 10.053]
24	(10.053, 10.43]
25	(10.43, 10.807]
26	(10.807, 11.184]

Continued on the next page...

Bin Number	Range
27	(11.184, 11.561]
28	(11.561, 11.939]
29	(11.939, 12.316]
30	(12.316, 12.693]
31	(12.693, 13.07]
32	(13.07, 13.447]
33	(13.447, 13.825]
34	(13.825, 14.202]
35	(14.202, 14.579]
36	(14.579, 14.956]
37	(14.956, 15.333]
38	(15.333, 15.711]
39	(15.711, 16.088]
40	(16.088, 16.465]
41	(16.465, 16.842]
42	(16.842, 17.219]
43	(17.219, 17.596]
44	(17.596, 17.974]
45	(17.974, 18.351]
46	(18.351, 18.728]
47	(18.728, 19.105]
48	(19.105, 19.482]
49	(19.482, 19.86]
50	(19.86, 20.237]
51	(20.237, 20.614]
52	(20.614, 20.991]
53	(20.991, 21.368]
54	(21.368, 21.746]
55	(21.746, 22.123]
56	(22.123, 22.5]
57	(22.5, 22.877]
58	(22.877, 23.254]
59	(23.254, 23.632]
60	(23.632, 24.009]
61	(24.009, 24.386]
62	(24.386, 24.763]
63	(24.763, 25.14]
64	(25.14, 25.518]
65	(25.518, 25.895]
66	(25.895, 26.272]
67	(26.272, 26.649]
68	(26.649, 27.026]
69	(27.026, 27.404]
70	(27.404, 27.781]
71	(27.781, 28.158]

Continued on the next page...

Bin Number	Range
72	(28.158, 28.535]
73	(28.535, 28.912]
74	(28.912, 29.289]
75	(29.289, 29.667]
76	(29.667, 30.044]
77	(30.044, 30.421]
78	(30.421, 30.798]
79	(30.798, 31.175]
80	(31.175, 31.553]
81	(31.553, 31.93]
82	(31.93, 32.307]
83	(32.307, 32.684]
84	(32.684, 33.061]
85	(33.061, 33.439]
86	(33.439, 33.816]
87	(33.816, 34.193]
88	(34.193, 34.57]
89	(34.57, 34.947]
90	(34.947, 35.325]
91	(35.325, 35.702]
92	(35.702, 36.079]
93	(36.079, 36.456]
94	(36.456, 36.833]
95	(36.833, 37.211]
96	(37.211, 37.588]
97	(37.588, 37.965]
98	(37.965, 38.342]
99	(38.342, 38.719]
100	(38.719, 39.096]
101	(39.096, 39.474]
102	(39.474, 39.851]
103	(39.851, 40.228]
104	(40.228, 40.605]
105	(40.605, 40.982]
106	(40.982, 41.36]
107	(41.36, 41.737]
108	(41.737, 42.114]
109	(42.114, 42.491]
110	(42.491, 42.868]
111	(42.868, 43.246]
112	(43.246, 43.623]
113	(43.623, 44.0]

References

[Abdelrahman *et al.* 2016] Hany Abdelrahman, Felix Berkenkamp, Jan Poland, and Andreas Krause. Bayesian optimization for maximum power point tracking in photovoltaic power plants. In *2016 European Control Conference (ECC)*, pages 2078–2083. IEEE, 2016.

[Ajmal *et al.* 2020] Aidha Muhammad Ajmal, Thanikanti Sudhakar Babu, Vigna K Ramachandaramurthy, Dalia Yousri, and Janaka B Ekanayake. Static and dynamic reconfiguration approaches for mitigation of partial shading influence in photovoltaic arrays. *Sustainable Energy Technologies and Assessments*, 40:100738, 2020.

[Bishop 2006] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, New York, 1st edition, 2006.

[Blei *et al.* 2017] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, 2017.

[Blundell *et al.* 2015] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural networks. In *Proceedings of the 32nd International Conference on Machine Learning (ICML-15)*, pages 1613–1622, 2015.

[Chen *et al.* 2020] X. Chen, Z. Li, and Y. Wang. Spatiotemporal graphical models for forecasting spatially distributed variables. *Journal of Machine Learning Research*, 21(152):1–35, 2020.

[Chiodo *et al.* 2015] Elio Chiodo, Davide Lauria, and Cosimo Pisani. Availability analysis of photovoltaic inverters in presence of uncertain data via bayesian approach. *IET Generation, Transmission & Distribution*, 9(13):1681–1687, 2015.

[Ciaramella *et al.* 2016] Angelo Ciaramella, Antonino Staiano, Guido Cervone, and Stefano Alessandrini. A bayesian-based neural network model for solar photovoltaic power forecasting. In Simone Bassis, Anna Esposito, Francesco Carlo Morabito, and Eros Pasero, editors, *Advances in Neural Networks*, pages 169–177, Cham, 2016. Springer International Publishing.

[Clerc and Kennedy 2002] Maurice Clerc and James Kennedy. The particle swarm—explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation*, 6(1):58–73, 2002.

[De Campos 2011] Cassio Polpo De Campos. Simulated annealing algorithms for the structure learning of bayesian networks. *Journal of Machine Learning Research*, 10(Mar):1659–1700, 2011.

[Derks and De Waal 2020] Iena Petronella Derks and Alta De Waal. A taxonomy of explainable bayesian networks. In *Artificial Intelligence Research: First Southern African Conference for AI Research, SACAIR 2020, Muldersdrift, South Africa, February 22-26, 2021, Proceedings 1*, pages 220–235. Springer, 2020.

[Duffie and Beckman 2013] John A. Duffie and William A. Beckman. *Solar Engineering of Thermal Processes*. John Wiley & Sons, 4th edition, 2013.

[Eberhart and Shi 2001] Russell Eberhart and Yuhui Shi. Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No. 01TH8546)*, volume 1, pages 81–86. IEEE, 2001.

[Freedman and Diaconis 1981] David Freedman and Persi Diaconis. On the histogram as a density estimator: L2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981.

[Friedman *et al.* 1997] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2-3):131–163, 1997.

[Gandomi *et al.* 2013] Amir H. Gandomi, Xin-She Yang, Siamak Talatahari, and Amir H. Alavi. *Metaheuristic Applications in Structures and Infrastructures*. Developments in Civil and Structural Engineering. Elsevier, Oxford, UK, 1st edition, 2013.

[Goldberg 1989] David E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, 1st edition, 1989.

[Heckerman *et al.* 1995] David Heckerman, Dan Geiger, and David M Chickering. Learning bayesian networks: The combination of knowledge and statistical data. *Machine learning*, 20(3):197–243, 1995.

[Heckerman 1998] David Heckerman. A tutorial on learning with bayesian networks. *Learning in graphical models*, pages 301–354, 1998.

[Hu *et al.* 2023] Pengcheng Hu, Abhisek Ukil, and Nirmal-Kumar C Nair. Photovoltaic maximum power point tracking based on bayesian optimization neural network. In *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, pages 1–6, 2023.

[International Energy Agency 2023] International Energy Agency. *World Energy Outlook 2023*. International Energy Agency, 2023.

[Jordan *et al.* 1999] Michael I. Jordan, Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. An introduction to variational methods for graphical models. In *Machine Learning*, pages 183–233. Springer, 1999.

[Kalogirou 2014] Soteris A Kalogirou. *Solar Energy Engineering: Processes and Systems*. Academic Press, 2nd edition, 2014.

[Kennedy and Eberhart 1995] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948. IEEE, 1995.

[Keyrouz 2018] Fakheredine Keyrouz. Enhanced bayesian based mppt controller for pv systems. *IEEE Power and Energy Technology Systems Journal*, 5(1):11–17, 2018.

[Khatib and Muhsen 2020] Tamer Khatib and Dhiaa Halboot Muhsen. Optimal sizing of standalone photovoltaic system using improved performance model and optimization algorithm. *Sustainability*, 12(6), 2020.

[Koller and Friedman 2009] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[Leicester *et al.* 2016] Philip A Leicester, Chris I Goodier, and Paul Rowley. Probabilistic evaluation of solar photovoltaic systems using bayesian networks: a discounted cash flow assessment. *Progress in photovoltaics: Research and applications*, 24(12):1592–1605, 2016.

[Li *et al.* 2018] Ting Li, Long Wu, and Yijia Cao. Bayesian network-based solar irradiance prediction considering spatial correlation. *Applied Energy*, 222:966–977, 2018.

[Masters 2013] Gilbert M. Masters. *Renewable and Efficient Electric Power Systems*. John Wiley & Sons, 2nd edition, 2013.

[Murphy 2002] Kevin P. Murphy. *Dynamic Bayesian Networks: Representation, Inference and Learning*. PhD thesis, University of California, Berkeley, 2002.

[Murphy 2012] Kevin P. Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[Najibi *et al.* 2021] Fatemeh Najibi, Dimitra Apostolopoulou, and Eduardo Alonso. Enhanced performance gaussian process regression for probabilistic short-term solar output forecast. *International Journal of Electrical Power & Energy Systems*, 130:106916, 2021.

[Neal 1995] Radford M. Neal. *Bayesian Learning for Neural Networks*, volume 118 of *Lecture Notes in Statistics*. Springer, New York, 1st edition, 1995.

[Neapolitan 2004] Richard E. Neapolitan. *Learning Bayesian Networks*. Pearson Prentice Hall, Upper Saddle River, NJ, 1st edition, 2004.

[Nojavan *et al.* 2016] Farnaz Nojavan, Song S Qian, and Craig A Stow. Comparative analysis of discretization methods in bayesian networks. *Environmental Modelling & Software*, 81:1–12, 2016.

[Poli *et al.* 2007] Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization: An overview. *Swarm Intelligence*, 1(1):33–57, 2007.

[Ribeiro *et al.* 2016] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144, 2016.

[Sangeetha *et al.* 2024] B. Sangeetha, K. Manjunatha, P. Thirusenthil Kumaran, et al. Performance optimization in photovoltaic systems: A review. *Archives of Computational Methods in Engineering*, 31:1507–1518, 2024.

[Schwarz 1978] Gideon Schwarz. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

[Sharma *et al.* 2023] Abhishek Sharma, Abhinav Sharma, Moshe Averbukh, Vibhu Jately, Shailendra Rajput, Brian Azzopardi, and Wei Hong Lim. Performance investigation of state-of-the-art metaheuristic techniques for parameter extraction of solar cells/module. *Scientific Reports*, 13, 07 2023.

[Shin *et al.* 2020] J. Shin, S. Park, et al. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *IEEE Access*, 8:64820–64830, 2020.

[Simplemaps 2024] Simplemaps. *South Africa Cities Database*. `https://simplemaps.com/data/za-cities`, 2024. Comprehensive database of 186 prominent cities in South Africa. Includes latitude, longitude, province, and other variables. MIT license. Available in CSV, Excel, and JSON formats.

[Singh and Kushwaha 2020] Chandershekhar Singh and Ajay Kushwaha. Machine learning-based solar energy forecasting: Implications on grid and power market. In *Intelligent Energy Management Technologies Conference*, pages 267–276. Springer, 2020.

[Smith *et al.* 2021] A. Smith, B. Johnson, and C. Brown. Deep learning-enhanced bayesian networks for time series forecasting. *Neural Networks*, 139:103–112, 2021.

[Sontake *et al.* 2021] V.C. Sontake, A.K. Tiwari, and V.R. Kalamkar. Experimental investigations on the seasonal performance variations of directly coupled solar photovoltaic water pumping system using centrifugal pump. *Environment, Development and Sustainability*, 23:8288–8306, 2021.

[Spirtes *et al.* 2000] Peter Spirtes, Clark Glymour, and Richard Scheines. Causation, prediction, and search. 2000.

[Storn and Price 1997] Rainer Storn and Kenneth Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[Tsamardinos *et al.* 2006] Ioannis Tsamardinos, Laura E Brown, and Constantin F Aliferis. The max-min hill-climbing bayesian network structure learning algorithm. *Machine learning*, 65(1):31–78, 2006.

[van der Meer *et al.* 2018] D.W. van der Meer, M. Shepero, A. Svensson, J. Widén, and J. Munkhammar. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using gaussian processes. *Applied Energy*, 213:195–207, 2018.

[Vasisht *et al.* 2016] M Shravanth Vasisht, J Srinivasan, and Sheela K Ramasesha. Performance of solar photovoltaic installations: Effect of seasonal variations. *Solar Energy*, 131:39–46, 2016.

[Wang *et al.* 2020] Wei Wang, Sen Li, Zheng Zhou, and Wanxing Sheng. A decision support system for solar energy integration in smart grids. *IEEE Transactions on Industrial Informatics*, 16(1):486–495, 2020.

[Zhang *et al.* 2020] Ruiyuan Zhang, Hui Ma, Wen Hua, Tapan Kumar Saha, and Xiaofang Zhou. Data-driven photovoltaic generation forecasting based on a bayesian network with spatial–temporal correlation analysis. *IEEE Transactions on Industrial Informatics*, 16(3):1635–1644, 2020.

[Zhang *et al.* 2024] Qianjin Zhang, Zhaorong Zhai, Mingxuan Mao, Shijing Wang, Siwei Sun, Dikui Mei, and Qi Hu. Control and intelligent optimization of a photovoltaic (pv) inverter system: A review. *Energies*, 17(7), 2024.