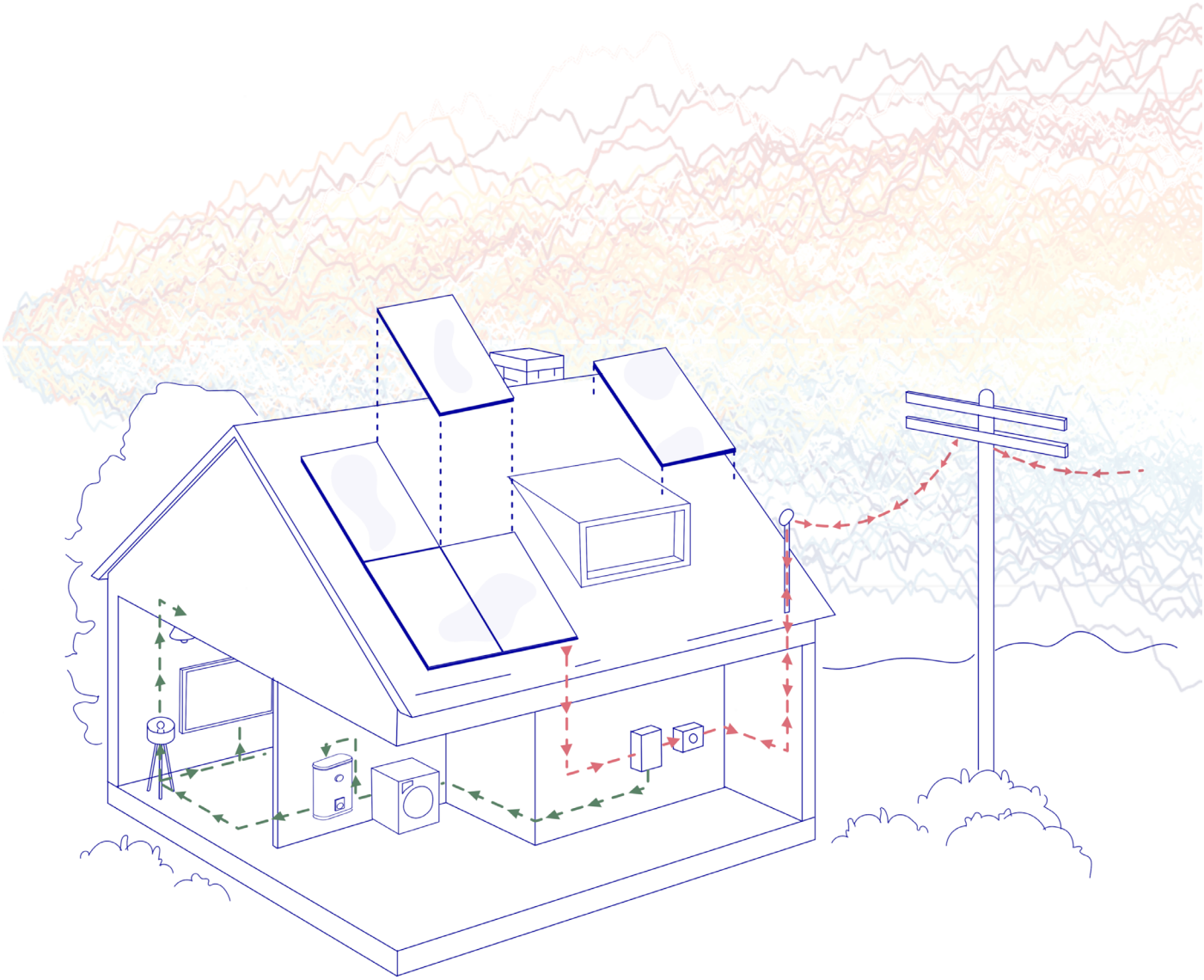
ReinforceRay

Optimal Long-Term Planning of Photovoltaic and Battery Storage Systems in Grid-Connected Residential Sector with Reinforcement Learning



TU Delft
Master Architecture, Urbanism and Building Sciences
Track Building Technology (Design Informatics + Building Energy Epidemiology)

AR3B012 GRADUATION STUDIO REPORT
2023/24

07.2024

Author: Jakub (Kuba) Wyszomirski - 5781078

1st Mentor: Dr. Michela Turrin
2nd Mentor: Dr. Charalapos Andriotis

Acknowledgments

I would like to thank my graduation mentors Michela Turrin and Charalampos Andriotis for their support and understanding. Additionally, I am grateful to Prateek Wahi, Dr. Regina Bokel, and Dr. Martin Tenpierik for their contributions to acquiring house load data ,and my colleagues and friends from the Building Technology track: Veronique, Jair, Alkis, Nefeli, Tahir, Shreya, Lara, Juan and more.

I dedicate this work to my mother Agata, my father Mariusz, and my girlfriend Weronika.

ReinforceRay

Optimal Long-Term Planning of Photovoltaic and Battery Storage Systems in Grid-Connected Residential Sector with Reinforcement Learning

TU Delft
Master Architecture, Urbanism and Building Sciences
Track Building Technology
AR3B012 GRADUATION STUDIO P4 REPORT
2023/24

07.2024

Jakub (Kuba) Wyszomirski - 5781078

Figure 01: TU Delft Logo (TU Delft, 2023)

Abstract

As the consumer electricity prices rise, European policymakers are increasingly focused on decarbonizing the power grid, which requires homeowners and local administrators to adopt renewable energy sources amidst a complex set of often conflicting objectives and constraints.

This paper introduces an innovative application of deep reinforcement learning (DRL) for long-term strategic planning of rooftop photovoltaic systems and battery energy storage within the residential sector, aiming to balance environmental and financial objectives considering the ever-evolving system condition and uncertainties inherent in the market.

The problem is modeled as a Markov Decision Process (MDP), facilitating sequential decision-making across 25 annual steps. The DRL environment incorporates a comprehensive set of variables identified through extensive literature review and market analysis. To account for their long-term dynamics, scenarios were simulated using appropriate stochastic and propabilistic processes for agent's training. A policy-based DRL agent is evaluated, exploring various residential and technological scenarios, including three single-family houses, different PV models and various optimisation scopes.

Moreover, a deployment workflow and a user interface are developed to support real-world decision-making applications. Furthermore, a separate DRL model is crafted to simulate battery management system's charging and discharging protocol.

The findings suggest that deep reinforcement learning offers a promising solution for addressing this complex problem. It offers enhanced flexibility in decision-making and helps mitigate investment risks.

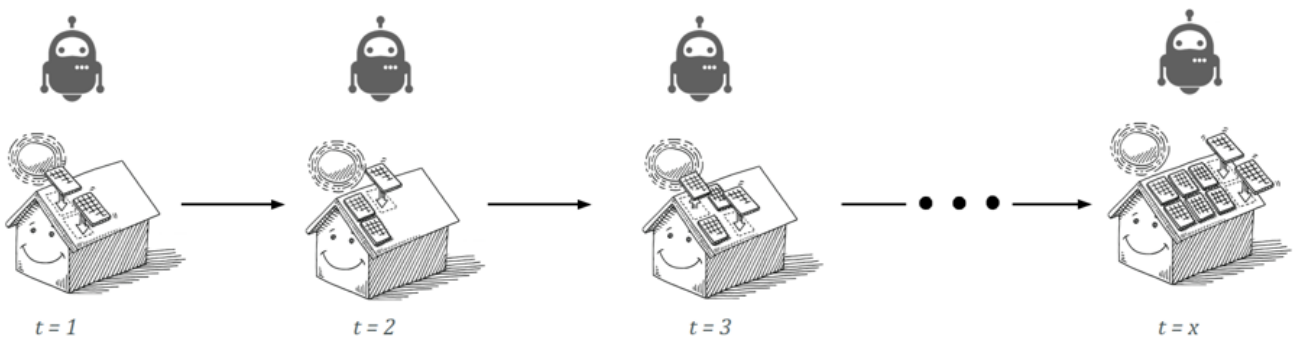


TABLE OF CONTENTS

Additionally asample code for this project is available on github:
https://github.com/kubawyszomirski/ReinforceRay

ACRONYMS

PV	Photovoltaics
GCRS	Grid-Connected Residential Sector
RL	Reinforcement Learning
BESS	Battery Energy Storage System
AC	Alternating Current
DC	Direct Current
OPEX	Operational Expenditure
CAPEX	Capital Expenditure
O&M	Operation and Maintnance
NPV	Net-Present Value
NCS	Net Carbon Savings
COE	Cost of Energy
PP	Payback Period
IRR	Internal Rate of Return
TMY	Typical Meteorogical Year
AOI	Angle of Incidence
BoS	Balance of System
MPD	Markov Decision Process
DRL	Deep Reinforcement Learning
ANN	Artificial Neural Network
RNN	Recurent Neural Network
LSTM	Long-Short Term Memory
CNN	Convolutional Neural Network
EG	Grid Emission Factor
MARL	Multi Agent Reinforcement Learning
DQN	Deep-Q-Network
PPO	Proximal Policy Optimisation
GBM	Geometric Brownian Motion
OU	Ornstein–Uhlenbeck Process
RS	Recommender System
OLS	Ordinary Least Squares
A2C	Advantage Actor Critic
LR	Learning Rate
LSTM	Long Short Term Memory
UI	User Interface
ROV	Real Option Valuation
API	Application Programming Interface

01. INTRODUCTION

01 INTRODUCTION

1.1 Context

1.1.1 Reinforcement Learning

Machine learning enhances decision-making by analysing vast amounts of data to identify patterns and insights that are often imperceptible to humans, thereby enabling more informed, efficient, and accurate choices. Among machine learning types, reinforcement learning (RL) stands out as particularly powerful for tasks where decision-making is sequential and the environment is uncertain. RL algorithms learn by interacting with their environment, receiving feedback in the form of rewards or penalties. It shows excellent decision-making capability in the absence of initial environment information (Sierla et al., 2022).

The implementation of RL in decision-making processes offers significant advantages. First, its versatility allows for its application across various fields through a combination of offline training and online execution, adeptly handling uncertainties in the process. Second, RL determines the best course of action through direct interaction with its environment, eliminating the need for pre-existing knowledge that can often be challenging to obtain. Third, compared to traditional optimization techniques, RL is more straightforward to apply in real-world situations (Xu et al., 2020).

Reinforcement learning has found extensive applications across multiple domains, including gaming, robotics, autonomous vehicle navigation, and financial trading, among others (Sutton & Barto, 2020).

In recent years there has been a upsurge of its applications to build environment. It demonstrates considerable promise in the field of building energy consumption and renewable energy sources adoption, particularly when balancing multiple objectives such as minimizing energy costs, reducing carbon emissions, and maximizing comfort (Fu et al., 2022).

1.1.2 Optimal Planning for Residential PV Systems

The last few years have seen a significant increase in energy prices across Europe, initially due to a rebound in energy demand following the relaxation of post-COVID lockdown measures, and subsequently after the Russian invasion of Ukraine. In the first half of 2023, average household electricity prices in the EU continued to rise compared to the same period in 2022, going from €25.3 per 100 kWh to €28.9 per 100 kWh (Eurostat, 2023).

To mitigate the impact of fluctuating energy prices and shortages, many homeowners are turning to renewable energy sources. Solar photovoltaic panels, often couple with a battery energy storage system (BESS) have become popular solution for generating electricity, reducing reliance on the grid, and even gaining energy independence within the grid-connected residential sector (GCRS). A PV system, when connected to the municipal network, provides power to meet the demand and send any surplus electricity back to the main grid (Khezri et al., 2022), the BESS on the other hand stores excess energy generated during peak sunlight hours and discharges it when the demand is high or sunlight is insufficient, ensuring a stable and continuous power supply.

Photovoltaic energy production is a crucial component of future energy solutions due to its non-polluting nature and high reliability (Schulte et al., 2022). The growing interest in PV installations can be attributed to various factors such as reduced costs, wider market accessibility, favourable government policies. The installation cost of both commercial PV systems and home batteries has decreased by 64% over the past ten years (Kurdi et al., 2022) .

Additionally, the introduction of microinverter technology enhances the scalability and adaptability of residential PV systems. Homeowners can conveniently enlarge their solar panel setups without having to ensure panel compatibility or overhaul the whole system (Morey et al., 2023).

Despite that, long-term planning of PV and BESS implementation and maintenance remains challenging. The upfront cost of installation can be significant. Additionally, the return on investment depends on various factors such as energy prices, solar irradiance in the location, and government incentives, which can vary over time (Jung et al., 2021). Installing PV panels without adequate planning may lead to surpassing the grid's capacity to absorb renewable energy (Kucuksari et al., 2014). Furthermore, the rates for feed-in-tariff and net-metering are declining in countries where rooftop PV systems are extensively adopted (Hayat et al., 2019). Homeowners who invest in current technology might find their systems becoming outdated sooner than expected, leading to a dilemma about whether to upgrade for better efficiency.

It has been demonstrated that uncertainty can influence decision-making, particularly when the investor has the opportunity to defer their investment. In this context, it's important to recognize that for a homeowner already

receiving power from the existing electrical grid, the decision to invest in solar PV systems is not mandatory but a choice (Moon & Baran, 2018).

1.2 State of the Art

The existing research on optimal PV + BESS planning in GCRS predominantly concentrates on single objective functions, usually evaluating techno-economic factors or system's reliability (Khezri et al., 2022). Studies often rely on fixed parameters and a specific time frame for evaluation without incorporating dynamic changes or variations that could occur over time. This approach is useful for initial assessment but may not fully capture the evolving nature of renewable energy systems and market conditions.

Linear programming has been used for municipal-scale planning of the rooftop PV deployment under budget constraints (Ren et al., 2023). Fan & Xia (2017) utilised non-linear integer programming for energy retrofitting including the implementation of PV panels. Integer programming however, relies on static models with fixed parameters.

Recent years have brought various machine learning based methods for GCRS PV implementation such as genetic algorithm (Khoury et al., 2015) or particle swarm optimization (Walker et al., 2020). However compared to RL those approaches focus more on immediate or short-term optimization for sequential decision-making problems without the same emphasis on long-term consequences.

Finally, RL has been used for the PV planning. Jung et al. (2021) utilised a geographic information system based RL to maximize the economic profit of the rooftop PV installation in various locations around Seul. This however did not take into account the electrical consumption of analysed buildings. On the other hand Liu et al. (2023) developed a RL algorithm for management of distributed energy resources in urban environments with diverse supply-demand profiles. Those, compared to the previously described methods, excel under volatile scenarios.

To author's knowledge to date no study describes the use of reinforcement learning for GCRS PV + BESS planning.

1.3 Problem Statement

In the mentioned context, optimal planning for residential PV + BESS must navigate a complex landscape of often conflicting and mutually exclusive variables and objectives. To date, the primary focus in optimizing the feasibility of residential PV installations is financial. However, those perspectives notably lack integration of other variables unrelated to economic profitability, such as environmental impact considerations.

Typically, optimization is approached as a singular event. However, there has been insufficient emphasis on longer-term analysis, which enables decision-makers not only to assess planning strategies but also to validate dynamic performance and implement system modifications or expansions throughout its lifespan, thereby achieving more practical outcomes.

Currently, existing methods prioritize "now-or-never" investment opportunities using discounted cash flow for optimization. Valuation is often tied to a specific moment in time, disregarding the potential to respond to evolving internal or external conditions that could delay the decision to install.

Formulating realistic plans can be challenging due to difficulties in accurately predicting time-dependent variables. Various studies employ deterministic optimization methods, thus offering investment choices that take into account a single operational scenario. However, the long-term accuracy of these assumptions or predictions can significantly influence the optimization outcome. Specifically, variables like electricity prices have demonstrated irregular fluctuations, complicating accurate forecasting (Jung et al., 2021). Unlike a deterministic model, a stochastic modeling approach accounts for the potential that various operational scenarios might occur in the future.

Reinforcement learning, with its robustness in handling uncertainties and complexities, is increasingly recognized for its effectiveness in long-term planning. Its performance under varying conditions remains strong (Fu et al., 2022). Despite this, the application of reinforcement learning, particularly for decision-making in building-integrated renewable energy sources, is still limited. In this context, **the problem lies in effectively utilizing the multi-objective optimization and sequential decision-making capabilities of reinforcement learning to navigate the complex, multidimensional spaces involved in long-term planning for rooftop PV systems in grid connected, residential scenario, especially when considering factors of divergent and often mutually exclusive natures, along with their inherent uncertainties.**

Figure 02: Research Landscape

1.4 Objectives

This research project aims to develop and evaluate a computational workflow for the optimal **real-option** planning of residential PV + BESS system adoption using RL, navigating the uncertainties of future scenarios. Ideally, this should result in a **creation of a trained RL model integrated into a recommender system, which provides commercial and private stakeholders and decision-makers, with a robust framework to guide and assess their investments in solar power and energy storage.**

Furthermore, the research seeks to identify and prioritize specific optimization objectives pertinent to this context. It will rigorously examine the variables influencing the optimal selection of actions within the defined action space, and assess their relative significance. This analysis will inform the reinforcement learning algorithm's reward mechanism, guiding its decision-making process within its environment. Various approaches to predicting the power generation of solar panel systems will be assessed. A critical aspect of this study involves generating future scenarios of energy prices for residential consumers and other relevant variables.

The research will involve a thorough evaluation and selection of specific reinforcement learning algorithms that are most suitable for addressing the problem at hand and assessment of PV system modeling methods.

The proposed framework shall be then evaluated on selected typologies of a typical Dutch single-family house in a residential setting, with an annual budget constraint for systems maintenance and installation.

1.5 Limitations

Often, the adoption of PV panels coupled with BESS is linked with complementary technologies like as turbines (Luo et al., 2019). However, in this research, such solutions are not considered. Additionally, the distribution gird related issues are not considered at it is assumed that derived results are technically feasible. Different phase connections between the inverters and the house loads are not considered; it is assumed that the connections between microinverters and household phases are optimal.

In this optimization process, the primary focus is exclusively on the capacity and type of the PV system and capacity of the BESS. Other elements like the tracking system, inverter type or capacity, and the tilt angle of the PV panels or different battery technologies are not taken into account.

Objectives other than the financial and environmental ones, which are frequently mentioned in the literature—such as aesthetics, production and operational capability, grid independence and coordination with and availability of power grids (He et al., 2017) —are not included in this analysis. Additionally, the analysis period is defined as 25 years as this is most commonly defined service life of PV modules in the most life cycle assessment (LCA) studies (Gerbinet et al., 2014).
This project does not claim to utilize the best models for capturing the uncertainties associated with the relevant variables. Instead, it focuses on what is feasible within the technical, logistical, and temporal constraints of this thesis. It is acknowledged that there may be superior methodologies that were not explored due to these limitations.

It is crucial to note that the developed approach is validated solely on synthetic data and the trained model is evaluated only within a controlled simulation setting and has not been tested against real-life scenarios or external environments, therefore it may not fully reflect the model's practical efficacy in real-world applications.

1.6 Research Questions

Realising the aforementioned theoretical background and problematic, the main research question and sub-questions of the thesis are stated as follows:

How can reinforcement learning based recommendation workflow be used for long-term planning and design of residential grid-connected PV and BESS under the uncertainty of future scenarios?

To be able to systematically reach the answer of the main research question, the following subordinate questions have been formulated:

How does the residential grid-tied photovoltaic system operate?

How can we evaluate the economic profitability and environmental benefits of rooftop PV systems and residential BESS?

Which variables to include in the optimisation process?

What constitutes the most appropriate model for forecasting the electricity yield of photovoltaic systems?

What constitutes the most appropriate model for simulating the operation of a home BESS?

How to generate future scenarios of the identified optimisation variables for model training?

Which reinforcement learning algorithm and in what configuration is most suitable for this problem?

What kind of action, observation spaces and reward function to consider?

How to deploy the trained model and how could it be applied for the end user?

1.7 Research Methodology

1.7.1 Introduction

The thesis begins with establishing its research framework, beginning with the underlying motivation that leads to the formulation of the research objectives and questions. These, in turn, shape the problem statement.

1.7.2 Literature Review

The second section encompasses the literature review process.

A number of key phrases are selected: reinforcement learning, building energy management, grid connected PV system planning, PV system modelling, BESS modelling, recommender system, machine learning deployment, electricity tariff forecasting, stochastic scenario generation etc. These keywords are searched in databases such as Google Scholar, Scopus, Science Direct, and Web of Science. Titles and abstracts are reviewed to identify relevant studies.

This phase is divided onto several steps:

A. Examination of residential grid-connected photovoltaic and battery systems. This serve to familiarise with the problem at hand, explore and discuss its key technical properties and components.

B. Identification of the factors influencing decision-making processes in financial and environmental contexts and an elaboration on their significance and impact.

C. Comprehension of the objective functions for optimisation as outlined in the literature and identifying the relevant design constraints.

D. Building on the foundational knowledge from sections A and B, this section will detail and discuss various methods for modeling photovoltaic panels and BESS within the framework of long-term planning.

E. Exploration of reinforcement learning, providing an overview of the reinforcement learning landscape and detailed descriptions of specific algorithms.

F. Building on the knowledge from step B concerning relevant variables and the specifics of reinforcement learning from the previous step, this section will explain and discuss the literature on creating stochastic scenarios using the identified independent variables for training the reinforcement learning model.

G. Finally reinforcement learning deployment considerations will be analysed.

1.7.3 Proposed Workflow

The third section will utilize findings from the literature review to create a detailed workflow plan, involving four crucial steps:

A. Simplified toy problem - serves as a testing and learning platform for the selected approach, facilitating the author's exploration and understanding.

B. Full-scale Development Part 1: This section encompasses all variables in their complexity, focusing exclusively on the photovoltaic (PV) system as the sole optimization variable.

C . Full-Scale Development Part 2 employs a methodology that builds on the previous step. Yet, both the PV system and the BESS are the focus of optimization.

D. Finally part D serves as a demostrative deployment with the desciprion of how the approach might be utilised for real decision-making.

This will be explained in more detail in Chapter 3 and the following chapters.

1.8 Relevance

1.8.1. Scientific Relevance

This thesis contributes to the practical understanding of applying RL for the optimization of residential PV and BESS systems and related decision-making. Scientifically, the research enriches the **field of applied energy informatics and home energy management systems** by providing a concrete example of how advanced computational techniques can address real-world problems in energy management. It ventures into dynamic optimization and real-options that accounts for uncertainty in long-term planning.

1.8.2. Societal Relevance

On the societal front, the research directly addresses the need for more efficient and environmentally friendly energy solutions at the household level. The development of a robust RL-based framework provides homeowners with a clear strategy to optimize their energy costs and reduce their carbon footprint. **The outcomes of this research could potentially lower the barriers to the adoption of PV and BESS by demystifying the economic and ecological trade-offs involved and by establishing a planning schedule.**

The trained model can be effectively used for strategic investment decisions PV and BESS, acknowledging the right but not the obligation to pursue or defer certain action based on evolving conditions. This methodology not only quantifies the financial benefits of strategic flexibility but also enhances the capability to manage risk more effectively in the face of uncertainty.

1.9 Ethics

With the rapid technological advancements in the fields of AI and machine learning over the past decade, particularly in generative AI, moral concerns have arisen regarding how AI might affect the roles of architects and engineers in the future. This thesis project's potential implementation could conflict with the interests of specialists, engineers, and businesses involved in home microgrid system design.

On the other hand, by employing reinforcement learning and other AI techniques, the project ensures that the methodologies and decision-making processes are transparent and justifiable. The incorporation of diverse scenarios and comprehensive testing under different conditions reflects a thorough approach to validate and justify the AI's decisions and recommendations, which is crucial for maintaining accountability

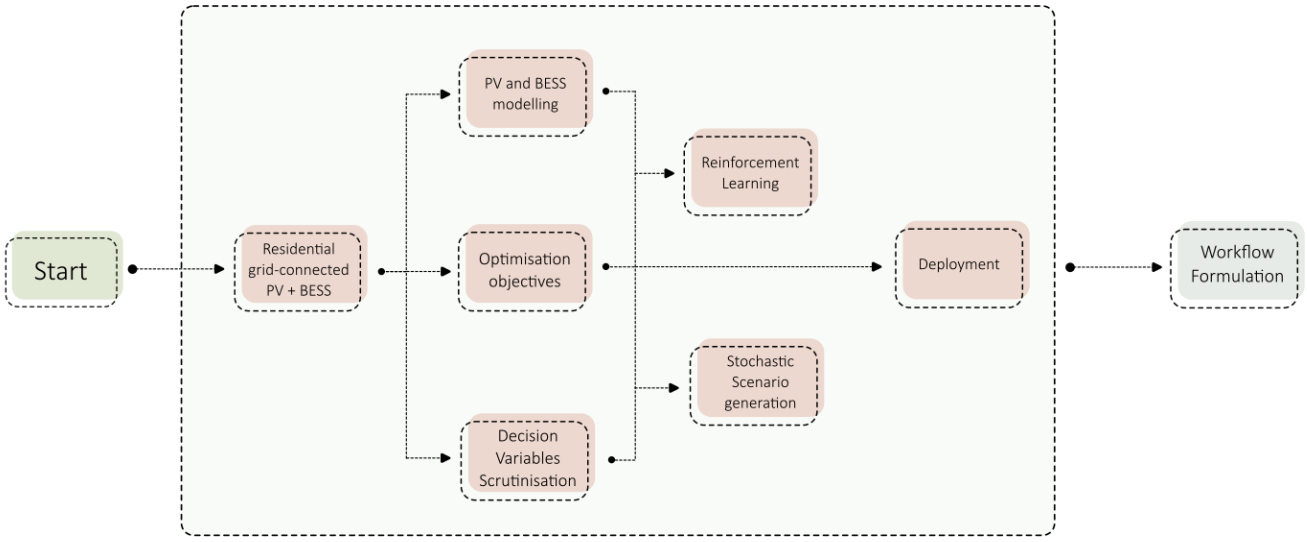


Figure 03: Literature Review Flowchart

02. LITERATURE REVIEW

2.1 Grid-Connected Residential System

2.1.1 Overview

Residential grid-connected PV systems are typically rated at less than 20 kW (Morey et al., 2023). Excess electricity generated by the system can be delivered back to the utility grid for utilization by other consumers. This reverse flow of power is monitored via a metering device that records the quantity of electricity redistributed. Typically, a photovoltaic system connected to the electrical grid eliminate the need for batteries because the grid serves as a backup energy source. Standard electrical grids supply alternating current (AC), whereas photovoltaic modules generate direct current (DC). To integrate PV panels with the grid, a DC/AC converter, commonly known as an inverter, is crucial (Lau et al., 2016). In schematic terms, this can be illustrated as in figure 04.

2.1.2 Solar Photovoltaic Modules

The PV system consists of multiple panels linked together in series to produce the necessary power, in accordance with the specified voltage and current ratings (Akın Taşcıkaraoğlu and Ozan Erdinç, 2019, pp. 49–62). The photovoltaic module is composed of the set of photovoltaic cells, these cells, primarily made of semiconductor materials, such as silicon are fabricated with a built-in electric field established by doping two adjacent layers of the semiconductor with different materials (Asif Hanif et al., 2021, pp. 226–243).

Upon absorption of sunlight, the energy of the photons is transferred to electrons in the atoms of the semiconductor material. Various types of silicon employed in the making of solar cells include single-crystal silicon, multi crystalline silicon, and amorphous silicon (Abo-Khalil et al., 2023). The effectiveness of a solar cell is gauged by its ability to transform solar radiation into electrical energy. A significant amount of the incoming solar energy is either reflected off or absorbed by the solar cell's surface, with only a minor portion being converted into electrical power (Asif Hanif et al., 2021, pp.226–243).

2.1.3 Inverter

The PV modules produce DC output, and the role of the DC-AC inverter is to transform the DC from the PV modules into AC, allowing it to be fed into the public distribution grid. Additionally, the inverter guarantees that the AC output is synchronized with the grid's voltage, frequency, and phase requirements. (Abo-Khalil et al., 2023). Additional components encompass a grid connection filter and an interaction monitor, responsible for tasks such as data measurement and anti-islanding (Kouro et al., 2015). Inverters enhance the performance of solar PV systems under variable conditions through Maximum Power Point Tracking (MPPT). This technology modifies the panel load to correspond with the fluctuating optimal power point, which is affected by alterations in sunlight intensity and ambient temperature .

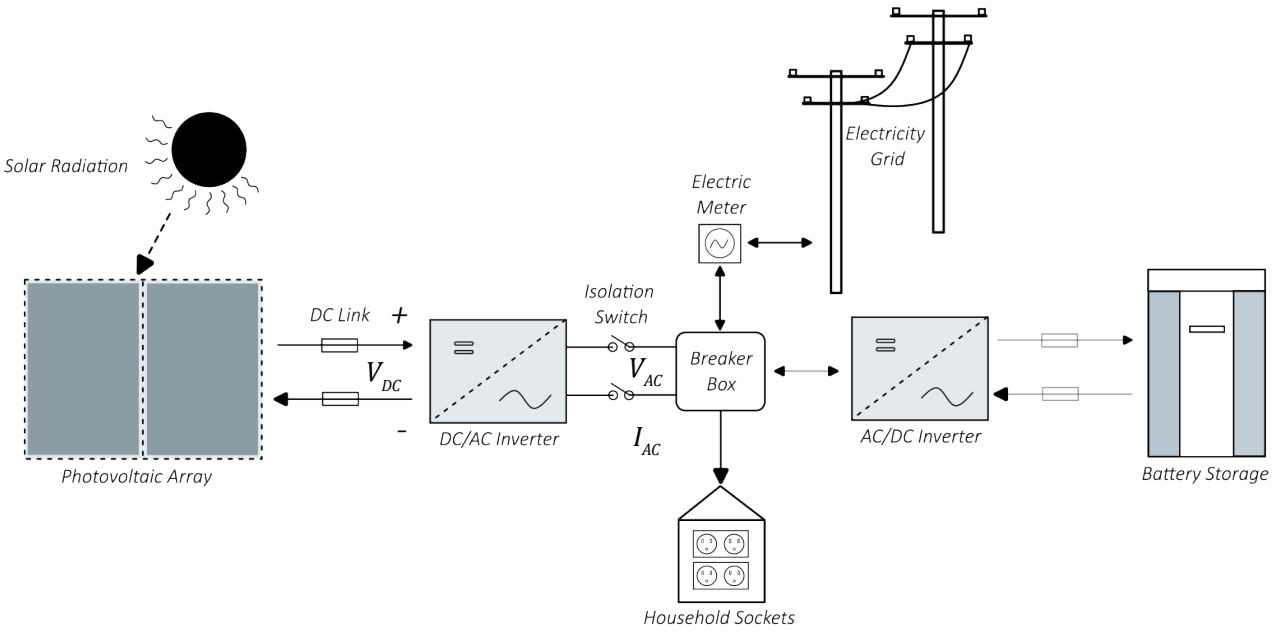


Figure 04: Residential Grid - Conected Photovoltaic System with BESS (author, 2024)

Inverters in grid-connected residential settings can be configured in several ways:

String inverter (fig. 5) - the most common option for grid-interfaced solar PV systems (Morey et al., 2023). String inverters have one centralized inverter connecting a series or ‘‘string” of solar panels, usually having power rating of 1 to 10 kW (Anzalchi and Sarwat, 2017).

Micro-inverter (fig. 6) - are integrated in each module. This allows each panel to function independently from its adjacent panels on the array, which means that the inefficiencies caused by variations in the power-voltage characteristics of individual modules are minimized. Micro-inverters are typically employed in systems with a peak power output up to 350 watts (Morey et al., 2023). They offer added flexibility by facilitating the easy expansion or reduction of modules, unlike string inverters which require a redesign of the entire system for such adjustments (Asif Hanif et al., 2021, pp. 252–254).

DC module inverter (fig. 7) - connects to the DC wires and AC output of solar panels. In contrast to micro-inverters, which convert DC to AC directly at the panel location, this system processes the DC electricity and conveys it to a centralized string inverter for conversion (Morey et al., 2023).

2.1.4 Building Load

The AC power from the inverter is then fed into the home’s electrical panel, also known as the breaker box. This panel, equipped with circuit breakers, distributes electricity to different circuits within the home (Hartner et al., 2017).

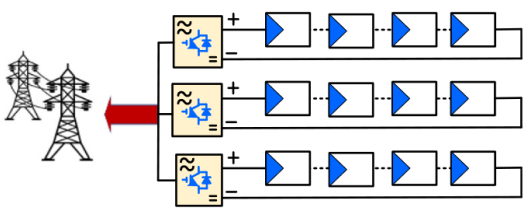


Figure 05: String Inverter Configuration (Morey et al., 2023)

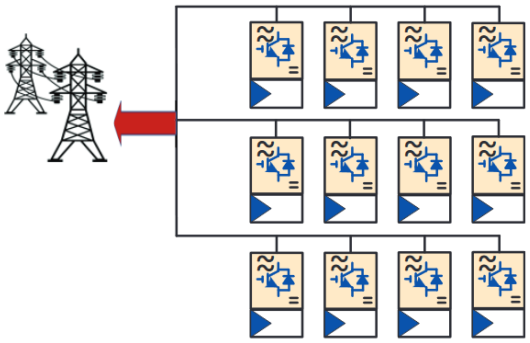


Figure 06: Micro-Inverter Configuration (Morey et al., 2023)

A standard daily load profile of a Dutch terraced house usually peaks in late afternoon hours, after dinnertime. Whereas the PV system energy production is the highest midday, when the average solar radiation peaks. Additionally, just like in other areas of the northern hemisphere, the household’s highest energy use nearly doubles in the winter, and the demand for energy in the morning also rises substantially. This leads to a notable difference between the PV energy production and usage, which is apparent in the total energy demand shown in, especially during the winter season (Klaassen et. al., 2015).

2.1.5 Home Battery Storage System

Home battery storage technologies have evolved significantly, with various options available based on chemical composition, connection type, and intended use. The primary types of home batteries include lead acid, lithium ion, nickel-cadmium, nickel-hydrogen, and more recently, sodium ion and salt water batteries. Each type presents distinct advantages and limitations in terms of cost, lifespan, energy density, environmental impact, and safety (Simpkins et al., 2016).

Lithium ion batteries dominate the residential and microgrid storage markets. They are favoured for their higher energy density, longer lifespan, and compactness

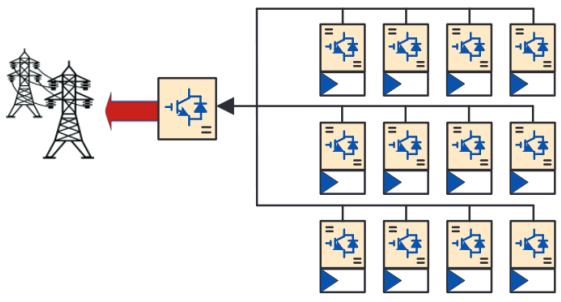


Figure 07: DC Module Inverter (Morey et al., 2023)

compared to lead acid batteries (Siewierski and Malarek, 2022). Various chemistries, including lithium nickel-manganese-cobalt (NMC), are used.

Lead acid batteries, traditionally used for automotive starters, have been adapted for home energy storage due to their low cost and well-understood recycling processes. However, they are bulkier, have a slower charge cycle, and are sensitive to high temperatures, which can shorten their lifespan (Chatzigeorgiou et al., 2024).

Emerging technologies like sodium ion batteries promise lower raw material costs and enhanced safety profiles compared to lithium ion. Salt water batteries, another innovative solution, are environmentally friendly due to the absence of heavy metals and toxic substances, though they suffer from lower energy and power capacities which restrict their use in high-demand applications (ibid.).

On the European market lithium ion batteries are the by far most popular choice for home energy storage systems primarily because of their superior energy density, which allows for a smaller physical footprint while providing greater energy storage capacity, and their longer cycle life, which offers greater overall value (Siewierski and Malarek, 2022).

The integration of these batteries into home energy systems can be either AC-coupled or DC-coupled. AC-coupled systems involve converting DC from solar panels to AC, which is then used directly in the home or converted back to DC for battery storage. This method, while versatile and compatible with existing solar setups, suffers from efficiency losses due to multiple conversions and typically has higher costs (Guan et al., 2015). DC-coupled systems store power directly from solar panels into batteries without the intermediate AC conversion, thus enhancing overall system efficiency and reducing component costs. These systems are particularly advantageous for new installations where system architecture can be optimized from the outset.

2.15 Conclusion

The integration of grid-connected photovoltaic systems into residential settings presents a multifaceted approach to sustainable electricity generation. The seamless interplay between the solar modules and the grid ensures a reliable flow of energy, balancing production with home consumption patterns.

Regarding the feasibility of the problem at hand it is best to utilize microinverters integrated within the solar panels. This setup provides greater flexibility, allowing for regular updates and adjustments to the system to maintain optimal performance.

Lithium-ion batteries are selected for optimisation. The dynamic optimization process described in this work requires the implementation of an AC-coupled system. This is because it allows for the potential addition of a battery at any point during the 25-year operational period without necessitating the replacement of the PV panel inverters.

2.2 Optimisation Variables and Objectives

This section analyses the factors that affect the long-term economic efficiency and environmental benefit of the GSPS found in the literature. This involves examining the various elements that influence the cost-effectiveness and financial performance of these systems.

2.2.1 Energy Balance and PV-related Variables

Energy balance is essentially the equation that describes the relationship between the energy output of the PV system and the residential energy consumption. The energy imported from the grid at time t, $E_{imp}(t)$, occurs when the consumption exceeds production, i.e., $E_{cons}(t) > E_{prod}(t)$. The total energy imported annually can be calculated as:

$$Total\ Energy\ Imported = \int_0^T max(E_{cons}(t) - E_{prod}(t), 0)dt \quad eq.\ 1$$

Similarly, the energy exported to the grid at time t, $E_{exp}(t)$, occurs when the production exceeds consumption, i.e., $E_{prod}(t) > E_{cons}(t)$.

$$Total\ Energy\ Exported = \int_0^T max(E_{prod}(t) - E_{cons}(t), 0)dt \quad eq.\ 2$$

Typically however this timestep is discretised to 15 minutes (Allouhi, 2020) or 1 hour (Hartner et al., 2017).

In the context of photovoltaic optimization, the solar radiation and the building load profile of the area under study is a critical element to consider from the perspective of energy demand (Al Garni et. al., 2018). Other factors including ambient temperature, air pollution or wind also influence the efficiency of the solar system. On the other hand, the technical data of a PV system such as the temperature coefficient, system efficiency, the tilt angle of the array, insolation under standard test conditions or the system losses are crucial in estimating system's energy output (Shiva Gorjian and Ashish Shukla, 2020, pp. 313–346). This will be explored in section 2.3 in more detail.

2.2.1.1 PV Efficency Improvement

Usually unacounted by literature is the fact that the photovoltaic cell efficency develops over time, improving the energy production capabilities of the PV systems (Vartiainen et al., 2020). According to Fraunhofer Institute for Solar Energy Systems (2023) the efficiency of crystalline silicon PV modules has been growing approximately 0.4% per annum over last 10 years.

Continuing this trend, it's projected that the average module efficiency could rise from 17.2% in 2018 to 30% by 2050. While single-junction silicon cells are bounded by a theoretical efficiency cap of around 30%, the use of multijunction cells can achieve significantly higher efficiencies (Vartiainen et al., 2020).

2.2.2 Finacial Variables

Typically, the overall expense of the PV is represented by combining the initial cost, ongoing maintenance expenses, and the cost of replacing the PV panels.

2.2.2.1. Capital Expenditure

Capital Expenditure (CAPEX) refers to the initial financial investment required to purchase and install the solar panel system. This includes various costs involved in the setup of a solar energy project, its key components being: the cost of PV panels, inverter, and balance of system (BoS) which encompases mounting and installation expenses such as wiring etc. (Vartiainen et al., 2019). On the European market the residential PV system CAPEX has been falling steadily over last two decades (European Commission, 2022), with module process hitting record-low of less than 0.15 €/Wp in September 2023 (Meban, 2023).

CAPEX is not just a list of costs but is typically expressed in a standardized format, often in €/Wp (Euros per Watt peak), as noted by Allouhi (2020), which is then multiplied by a coefficient considering the base size of the system (Erdinc et al., 2015).

Literature usually consideres CAPEX to be constant (Mondol et. al., 2009), or its original value at the start of the the optimisation period is adjusted annually with an interest or discount rate (Hartner et al., 2017). For example, in order to determine the annuity's worth based on the interest rate and the duration of the payment Khoury et al., (2015) has multiplied the base CAPEX by a capital recovery factor. A government subsidy covering a fraction of the CAPEX is also sometimes taken into consideration (Zhou et al., 2018).

Sometimes more than one pricing scenarios are considered (Liu et al., 2012). Far less often future CAPEX is estimated stochastically considering among other factors past data (Jung et al., 2021). This approach will be further explored in the section 2.5.

2.2.2.2. Operational Expenditure and Replacement

Operational expenditure (OPEX) refers to the annual,

Operational expenditure (OPEX) refers to the annual, ongoing costs associated with operating and maintaining the PV system over its lifetime. It's main component is operation and maintenance (O&M), which includes cleaning the panels, checking and repairing electrical connections, and replacing components (Vartiainen et al., 2019). Not all studies take OPEX under consideration, or assumed that its influence is negligible (Nafeh, 2009).

OPEX can be expressed as an independent variable (Al Garni et. al., 2018), where the initial value of OPEX is gathered from the empirical data and then adjusted annually considering the interest rate. For example, Kornelakis and Marinakis (2010) have considered the PV panel maintenance costs and the inverter repair adjusted for both interest rate and inflation. Cervantes and Choobineh (2018) assumed different scenarios for the maintenance and labour costs depending on the location of the system under optimisation. Alternatively OPEX is calculated as percentage of the capital cost, usually as 1% (Khoury et al., 2015) or 2% (Hartner et al., 2017) of the CAPEX value of the entire system per year.

Most analysed studies do not consider replacement costs or assume they are equal to the CAPEX per panel (Fan and Xia, 2017). Others only weight the costs of the replacement of the inverter after a certain time period (Javeed et al., 2021).

2.2.2.3. Energy pricing and Net-metering

The structure of electricity pricing plays a pivotal role in the decision-making process for installing PV systems. In the Netherlands various types of energy contracts are in place (fixed, time-of-use and real time pricing) (van de Wetering, 2023).

With fixed rates, the price for importing power remains constant as agreed with the supply company. Under time-of-use pricing, electricity costs are typically segmented into distinct periods throughout the day, higher rates are applied during the day's peak demand hours. Al Garni et al., (2018) considered three distinct daily thresholds – peak, shoulder and off peak. In real time pricing, the electricity rate is adjusted every hour in a dynamic manner (Zhou et al., 2018). Yet, the majority of studies regarding the financial optimisation for GCPS assume fixed contracts billed monthly or weekly from the consumers (Khezri et al., 2022).

All reviewed papers have taken into account a form of compensation for returning surplus energy to the power grid. This typically involves a feed-in tariff (FIT) or a net metering arrangement. Analogous to the energy tariff, the FIT can be structured either as a fixed rate or variably, based on the time when the energy is supplied to the grid (Javeed et al., 2021). In the Netherlands, net-metering ("Salderingsregeling") has been a legislative framework, which mandates companies to offset the quantity of electricity fed back into the grid by a household against its consumption from the grid. This process effectively means that households are monthly billed only for their net electricity usage (Butenko, 2016).

2.2.2.4. Interest Rate

When optimising for long-term economic viability of a PV system most authors include the annual interest rate into the calculations. It usually is considered in the discount factor which adjusts future money to present value, considering that money available at the present time is worth more than the same amount in the future due to its potential earning capacity. This is in order to account for inflation, future investment oportunity, risk, uncertainty and a general people's preference for immediate consumption. Especially when long-term projects like PV installations is considered, where costs and benefits occur at different times (Khezri et al., 2022).

2.2.3 Environmental Variables

2.2.3.1 Grid Emission Factor

When evaluating the PV system in terms of the environmental benefit the grid emission factor (EG) is the most important variable. Compared to the amount of equivalent CO2 offset by a PV array, the emissions from PV system production can be deemed negligible (Liu et al., 2012). In literature it is expressed as the amount of CO_2 equivalent emitted per unit of electricity generated in kilograms of CO_2 per kilowatt-hour (Scarlat et. al., 2022). Allouhi, (2020) for the analysed time period used linear approximation to estimate the EG between the present value and a future predicted value obtained from the literature.

2.2.3.2 Embodied Carbon

Determining the carbon footprint of an on-roof photovoltaic installation and home BESS through a life cycle assessment is a complex task. This process needs to take into account various factors, including the energy and materials required for the production of photovoltaic panels, their transportation to the installation site, the energy consumed during the installation and production process, the electricity grid's factor, the operational

phase of the system over its lifespan, its end of life treatment and disposal. The complexity of this process and its uncertainties could be the reason that none of the examined studies take this factor into account for optimization.

2.2.4 Constraints

Apart from the design variables, the literature describes several design constraints that limit the planning of the PV system. Those include spatial constraint or the rooftop availability, Kornelakis and Marinakis, (2010) have limited the maximum number of rows of PV modules and considered their arrangement regarding the inverter configuration and shading. Fan and Xia, (2017) tested different budget constraints for the total spending on the PV implementation. System's resilience to extreme events can also be tested, where random power outages are modelled (Khoury et al., 2015). Javeed et al., (2021) and Cervantes and Choobineh, (2018) imposed maximum daily energy export caps to the utility grid due to local policies.

2.2.5 Decision Variables

The optimisation variable of the majority of the literature is the total capacity of the photovoltaic system. This is expressed either as the number of the photovoltaic modules (Khoury et al., 2015) or their combined power output in kW (Liu et al., 2012). Other studies considered a fixed, small number of the possible PV array sizes (Zhou et al. 2018) and (Lau et al., 2016). Usually only one type of PV module is selected to the optimisation.

The system capacity is usually optimised "statically" i.e. one specific configuration is selected for the entire lifespan of the project. Only Fan and Xia, (2017) have considered a scenario when a number of panels is added or discarded from the system each year.

2.2.6 Optimisation Objectives

In the realm of optimizing GCPS, objective functions is the most important parameter in guiding the design and operational strategies. Various optimization algorithms are utilized to either maximize or minimize objective functions. These problems in optimization can be formulated with either a single or multiple objective functions, depending on the complexity and requirements of the task.

Most of the revied literature optimise the GCPS in order to maximise the financial profit from the system implementation. These include net present value (NPV), cost of electricity (COE), payback period (PP), the internal rate of return (IRR) or other tailored formulas. The former two are by far most common. System NPV signifies its total value, adjusted for discounting, accounting for all electricity cost savings generated by the solar PV system, minus the initial investment costs associated with the technology adoption. COE is determined by dividing the total NPV by the annuity factor, that converts a series of future cash flows into a single present value. NPV is given by (Al Garni et. al., 2018):

$$NPV = \sum_{t=1}^{n} \frac{R_t - C_t}{(1+r)^t} \qquad \textit{eq. 3}$$

where:
R_t is the revenue at time t
C_t is the cost at time t
n is the number of periods
r is the interest rate

Some studies (Allouhi, 2020) also include an environmental objective such as the cumulative environmental benefit (CEB). This is calculated as the CO_2 grid emissions which are neutralised by the project electricity production. A comprehensive overview of the optimisation objectives, constraints and independent variables found in the literature has been given in the table 2.

While the literature reveals a diversity in optimization approaches, it focuses predominantly on financial objective. Although, these are crucial for evaluating the economic viability, it becomes clear that the emphasis often overlooks other important aspects such as environmental impacts.

2.2.7 Conclusion

Operational and capital expenditures, along with energy pricing, net-metering policies, and various incentives, are scrutinized for their role in shaping the feasibility and profitability of solar investments. Constraints such as spatial limitations and system resilience; decision variables as system capacity; and optimization objectives focused mainly on financial gains are thoroughly examined. It becomes evident that while economic viability remains the primary focus, the incorporation of environmental benefits into the optimization models is gaining traction. All of the variables described in this subchapter are pivotal to the optimal PV system planning and all shall be taken into account in the development of the RL model.

Decision Variables	Objective	Constraints	Electricity Tariff	Reference
PV capacity and solar tracking system	Financial - NPV	Spatial	Time-of-use	Al Garni et al., (2018)
PV capacity, inverter capacity, PV inclination	Financial – COE	-	Fixed	Mondol et al., (2009)
PV capacity	Financial - NPV	-	Fixed	Lau et al., (2016)
PV capacity	Financial - NPV	Power outages	Fixed	Khoury et al., (2015)
PV capacity and inclination, PV distribution among inverters	Financial - NPV	Spatial	Fixed	Kornelakis and Marinakis, (2010)
PV capacity and inclination	Financial – ROI	-	Fixed	Liu et al., (2012)
PV capacity	Financial - other	-	Fixed	Khanfara, (2018)
PV capacity	Financial - IRR	Policy	Fixed	Hartner et al., (2017)
PV capacity	Financial – NPV, PP	Spatial and Budget	Fixed	Fan and Xia, (2017)
PV capacity	Financial - PP	-	Fixed	Abo-Khalil et al., (2023)
PV capacity	Financial - COE	-	Fixed	Nafeh, (2009)
PV capacity	Financial – COE, Autonomy	Policy	Time-of-use	Ramli et. al., (2018)
PV capacity	Financial – COE, Environmental - CEB	Technical	Fixed	Allouhi, (2020)
PV capacity	Financial – sensitivity analysis	Technical	Fixed	Erdinc et al., (2015)
PV capacity	Financial – other	Technical	Fixed, Time-of- use, Real-time pricing	Zhou et al. (2018)
PV capacity	Financial - other	Policy	Time-of-use	Cervantes and Choobineh, (2018)
PV capacity	Financial – NPV, COE	Policy and spatial	Fixed, Time-of-use	Javeed et al., (2021)

Table 02: Overview of Literature Regarding the Optimal Planning of a GCPS for Residential Scenarios

2.3 PV and BES System Modelling

The underlying physics of a solar cell are quite sophisticated. The modelling of a cell's reaction to light and heat varies greatly, from simple high-level estimates requiring only a few coefficients to detailed electrical circuit models that need comprehensive cell data, and up to complete physics-based semiconductor models (Jo & Lee, 2019). The atmospheric conditions for the calculations are usually based on the typical meteorological year (TMY) weather data (E3P, 2016).

2.3.1 Solar Radiation

Estimating the intensity of solar radiation reaching the Earth's surface encompases the disaggregation of global horizontal irradiance into direct and diffuse components and the transposition of these irradiances onto a plane of fixed inclination to reflect the orientation of photovoltaic modules (Kazem & Yousif, 2017). The aggregate solar radiation on tilted surfaces is a composite of several components: diffuse horizontal radiation, direct beam radiation and reflected radiation (Perez et al., 1990). This can be computed as:

$$G_T = G_{h,b} \times \cos AOI + G_{h,d}\left(\frac{1+cos\theta_T}{2}\right) + G_{h,p}\left(\frac{1+cos\theta_T}{2}\right)$$

eq. 4

where:
$G_{h,b}$ is the direct beam radiation
$G_{h,d}$ is the diffuse horizontal radiation
$G_{h,p}$ is the reflected radiation
AOI is the angle of incident sunlight on a tilted plane
θ_T is the tilted angle of inclination relative to the ground surface

$G_{h,b}$ is sunlight that travels in a straight line from the sun to the Earth's surface. $G_{h,d}$ has been scattered by molecules and particles in the atmosphere. Reflected radiation, is sunlight that has hit the Earth's surface and been reflected back into the atmosphere (Perpetuo e Oliveira et al., 2019).

AOI between the Sun's rays and the PV array can be determined (Hosseini et al., 2018):

$$AOI = cos^{-1}(\cos\beta\ \cos(A-\varphi)\, sin\theta_T + sin\beta\ cos\,\theta_T)$$

eq. 5

where:
φ is the azimuthal deviation of the PV module
A is the sun azimuth angle
β is the sun altitude angle.

2.3.2 Cell Temperature

Rising temperature of PV cells results in diminished efficiency due to increased carrier recombination and reduced bandgap energy. Therefore, it's important to accurately determine the cell temperature to predict the power a PV system can generate. The equation for cell temperature T_C taking into account wind speed and the mounting configuration is defined as (Faiman, 2008):

$$T_c = T_{ambient} + \frac{\alpha \times G_T(1-\eta_m)}{U_c \times U_v \times WS}$$

eq. 6

where:
$T_{ambient}$ is the local ambient temperature,
$T_{c,STC}$ is the PV temperature under standard test conditions [25°C],
α is the PV cell absorption coefficient,
U_c is the heat loss factor coefficient dependent on the module construction,
U_v is the combined heat loss factor influenced by wind,
WS is the wind speed in m/s.

2.3.3 Energy Output

In literature on optimal solar PV planning in GCRS different power output models are used. Mondol et al. (2009) modelled grid-connected PV system using the TRNSYS, graphically based software environment, where a "four-parameter" equivalent circuit is employed to model a single crystalline PV module. This model is characterized by the module's photocurrent under STC, its diode reverse saturation current at STC, an empirical factor for fitting the PV curve of the diode, and the module's series resistance. This is called a **single-diode model**. TRNSYS is also used by Kornelakis & Marinakis (2010) for the optimisation of the PV surface inclination for the PVGCS energy production. Al Garni et al. (2018) and Lau et al. (2016) carried out the simulations using HOMER for design of off-grid and on-grid power systems, commonly used for renewable energy sources evaluation and sensitivity analysis.

Other computational tools like Sunny Design (Khanfara et al., 2018), RETScreen, PVSyst, Hybrid2 (Connolly et al., 2010) or self-developed sotwares (Hartner et al., 2017) are also among the most frequently reported in the literature.

Despite the complexity of solar PV systems other authors have utilised **simplified formula-based approaches**. Liu et al. (2012) calculates the actual power output of a PV array based on its rated capacity and the solar irradiation it receives. Similarly, Jung et al. (2021) and Ogunjuyigbe

et al. (2016) modelled PV energy output as a product of area, efficiency of solar panels and the global solar radiation.

More comprehensive approach, than a simple linear formula factors in effect of temperature on the efficiency of the PV panel, as well as the variance in solar irradiation. It can be called a **single-point efficiency** with temperature correction model (Theristis et al., 2018, pp. 671–706) (eq. 7). This were utilised by Fan & Xia (2017) and Liu et al. (2023) taking into account that the electrical properties of PV systems vary from the parameters measured under STC. Even more detailed analytical approach suitable for precise calculations was developed by Khoury et al. (2015) incorporating detailed physical and electrical characteristics of the PV cells.

$$P_{PV} = Y_{PV} \times \left(\frac{G_T}{G_{T,STC}}\right)\{1 + \alpha_P(T_c - T_{c,STC})\} \quad \textit{eq. 7}$$

where:
P_{PV} is the power generated from the PV panel
Y_{PV} is the rated capacity of the PV panel [kW]
$G_{T,STC}$ is the incident radiation in STC [kW/m2]
G_T is the solar radiation on the PV [kW/m2]
α_P is the temperature coefficient [%/°C]
T_c is the cell temperature [°C]
$T_{c,STC}$ is the PV temperature under STC [25°C]

Contrary to the previously mentioned methods, **Data-Driven Regression Modelling**, as detailed by Allouhi (2020), employs statistical methods like Multiple Linear Regression, utilizing historical data to establish model coefficients. This is particularly suitable for situations where historical data is accessible and environmental conditions are relatively stable.

Not mentioned in the GCPS literature is the recently released Python library pvlib developed at Sandia National Laboratories, integrated with other Python-based data analysis tools (Holmgren et al., 2015) is able to calulcate the PV power output using both the single-diode and single-point models under various PV system configurations and environmental conditions.

In general the appropriateness of a modelling approach depends on its intended use – basic estimations can often yield precise results, but might not be suitable for making financial decisions. Conversely, detailed physics-based models of semiconductors are typically only useful in research environments, not in practical, real-world applications (Connolly et al., 2010).

2.3.3.1 Single-diode Modelling

The single diode model for representing the electrical behaviour of a photovoltaic cell, which has been proven to accurately approximate the PV performance (Theristis et al., 2018, pp. 671–706). This model is based on the equivalent circuit of a diode (fig. 08), which reflects the PV cell's characteristics under various conditions of irradiance and temperature. It captures the current-voltage (I-V) relationship of a PV device and generates the I-V curve (fig. 10), which plots the current that a solar cell generates against the voltage it produces. The current I flowing through the cell is represented as (Wenham et al., 2011):

$$I = I_L - I_0\left[exp\left(\frac{V+I\times R_S}{n\times N_s\times V_{th}}\right) - 1\right] - \frac{V+I\times R_S}{R_{sh}} \quad \textit{eq. 8}$$

where:
I_L is the light-generated current, produced by the cell due to light exposure,
I_0 is the diode saturation current, flowing through the diode when the voltage is below the threshold,
R_S is the series resistance, representing the resistive losses of the cell,
R_{Sh} is the shunt resistance to the current that leaks across the p-n junction,
n models the deviation from ideal p-n junction behaviour,
N_S is the number of cells connected in series within each module,
V_{th} is the thermal voltage depending on the temperature and the physical properties of the materials in the cell.

The five parameter values for this equation cannot be found in manufacturer's datasheets. Their estimation procedure is described in Appendix A. The voltage V across a PV cell can be determined using the Lambert W function. This is given by :

$$V = -I \times R_s - I \times R_{sh} - nV_{th} \times LambertW\left(\frac{I_0 R_{sh} e^{\left(\frac{R_{sh}(-I+I_{ph}+I_0)}{nV_{th}}\right)}}{nV_{th}}\right)$$

eq. 9

where:
I is the actual current through the PV cell,
$LambertW$ is the lambert function defined as the function that satisfies the equation $x=W(x)e^{(W(x))}$

Power is simply the product of current I and voltage V at any given point on the I-V curve. Hence:

$$P_{DC} = I \times V \quad \textit{eq. 10}$$

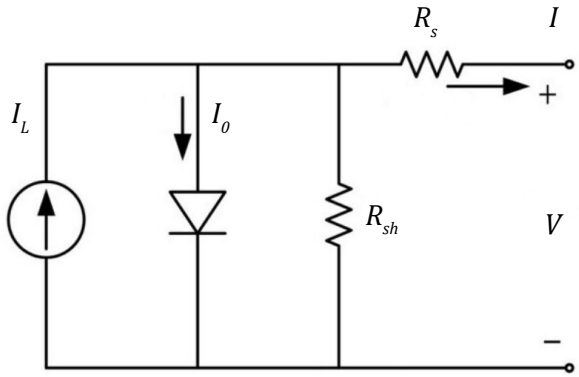


Figure 09: A Single-diode Model (author, 2024)

2.3.4 Soiling Losses

Soiling losses refer to the reduction of panel's efficiency due to the accumulation of dirt, dust, sand, or other particles on their surface. Typically, the industry standard estimates for annual soiling losses in photovoltaic systems range between 1% and 4% (Theristis et al., 2018, pp. 671–706). In the Netherlands' climate these are not as significant due to frequent rainfall. Nevertheless various empirical models have been developed to calculate the soiling loss as a function of environmental factors. Kimber Soiling model (Kimber et al., 2006) assumes that the accumulation of dirt occurs at a steady pace until it is washed away by rainfall.

2.3.5 Inverter Losses

Inverter losses in a PV system refer to the energy that is not converted from DC to AC due its inefficiency. The inverter's efficiency typically varies with the load, and it's often lower at very low or high power outputs. Predicting inverter losses in a PV system involves several factors including the inverter's efficiency curve and the expected load profile. Usually the AC-DC conversion account for around 3% power losses (Kazem & Yousif, 2017), this figure however is usually lower for microinverters. Important to note that very few studies

Method	Implementation Complexity	Accuracy
Simplified Formula-based	Easy	Poor
Single-point Efficency	Easy	Mediocre
Single-diode	Difficult	High

Table 03: Basic Comparison Between Described PV Modelling Methods

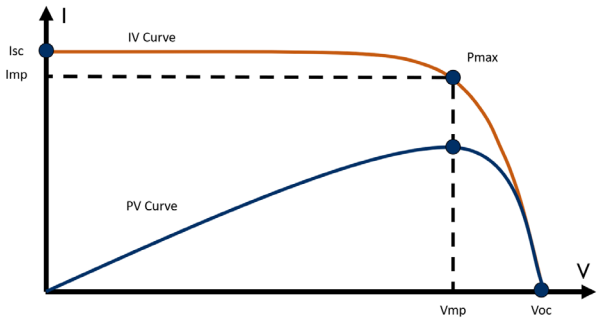


Figure 10: The I-V Curve (orange) and PV power production curve (blue)

regarding the PV planning for GCPS consider any losses when calculating the PV energy yield, assuming that the array DC output as conclusive.

Inverter efficiency is difficult to estimate and usually is evaluated based on empirical tests and observations (Perpetuo e Oliveira et al., 2019). The Sandia National Laboratories model for the AC power output from the inverter is given as (King et al., 2007):

$$P_{AC} = \left[\frac{P_{AC,0}}{A-B} - C(A-B)\right] \times (P_{DC} - B) - C(P_{DC} - B)^2 \quad \textit{eq. 10}$$

where the coefficients A, B, and C are defined as:

$$A = P_{DC,0} \times \left[1 + C_1(V_{DC} - V_{DC,0})\right] \quad \textit{eq. 11}$$

$$B = P_{DC,0} \times \left[1 + C_2(V_{DC} - V_{DC,0})\right] \quad \textit{eq. 12}$$

$$C = C_0 \times \quad \left[1 + C_3(V_{DC} - V_{DC,0})\right] \quad \textit{eq. 13}$$

where:
$P_{AC,0}$ is the maximum AC power,
$P_{DC,0}$ is the DC-power level at which the max. AC-power rating is achieved,
$P_{S,0}$ is the power consumption during operation,
V_{DC} is the PV maximum power voltage,
$V_{DC,0}$ is the DC-voltage level at which the max. AC-power rating is achieved,
C_0 is the parameter defining the parabolic relationship between AC and DC power at reference conditions.

2.3.6 Cabling Losses

Cabling losses in solar PV systems refer to the power losses that occur in the DC cables connecting the solar panels to the inverter, and in the AC cables connecting the inverter to the load or grid. According to Hashemi et al. (2021) these losses are influenced by factors like the potential difference over DC wiring. The cabling loss can be calculated as a function of resistance within the cables and voltage between cable ends (Ekici and Kopru, 2016).

2.3.7 PV Efficiency Loss

PV efficiency loss refers to the rate at which solar panels lose their efficiency over time. Solar panels, like all electronic devices, experience wear and tear, and their ability to convert sunlight into electricity diminishes gradually (Olczak, 2023). It's typically expressed as a percentage loss in efficiency per year. Literature indicates that the degradation rates vary between 0.8% and 4.9% per year, also varying over time since the original installation (Rahman et al., 2023).

There are many variables that influence the long-term efficiency loss rate starting from the manufacturing process, proper maintenance and installation. High temperature and temperature amplitude causes thermal stresses inside the panels causing delamination of the components. Moisture and air humidity leads to corrosion and electrical leakages (Jordan & Kurtz, 2011).

Regarding the studies on GCPS the efficiency loss is rarely taken into account, yet it should be considered in the long-term planning scenario, since lower energy output can affect the return on investment for solar installations, extending the payback period.

The literature outlines several methods for modeling the efficiency loss of photovoltaic panels. Firstly, a degradation rate method employs a percentage reduction in efficiency, which may be either a fixed value derived from measurements, manufacturer data, or cited literature, or it may be modeled stochastically using techniques such as gamma distributions (Jung et al., 2021), Monte Carlo simulations or sampled from a relevant distribution (Lai et al., 2023). Additionally, temperature coefficient modeling adjusts the panel's efficiency based on its temperature coefficient (Paudyal and Imenes, 2021). More advanced methods, such as physical degradation modeling, integrate physical degradation factors like microcracks, delamination, and corrosion using predictive algorithms (Guerra et al., 2023). Furthermore, electrical circuit models are used to simulate the evolution of key solar cell parameters (outlined in section 2.3.3.1), which influence efficiency over time (Arar et al., 2019).

2.3.8 PV Service Life

Apart from the efficiency loss the output from a PV array inevitably decreases due to the declining number of operational solar panels within the system as time progresses (Fan & Xia, 2017). The mean service life of PV modules given by the manufacturers usually ranges between 20 to 25 years. In reality, the lifespan of photovoltaic systems is not fixed and can varies. Therefore, it's important to accurately model the uncertainty in their service life for effective long-term planning of PV systems.

Ma et al., (2022) have built empirical model on the basis of 4 types of degradation patterns: hygrothermal, UV radiation temperature and humidity. Fan et al., (2018) have applied gamma process with an exponential transformation to PV modules to obtain the estimated lifetime.

A simpler approach has been taken by Laronde et al., (2011) survival over time has been represented using the Weibull distribution, a widely-accepted approach for assessing reliability and lifespan of technology in various applications. The effectiveness of this approach is confirmed by Kuitche (2010) who compared Weibull distribution modelling to gamma and exponential distributions.

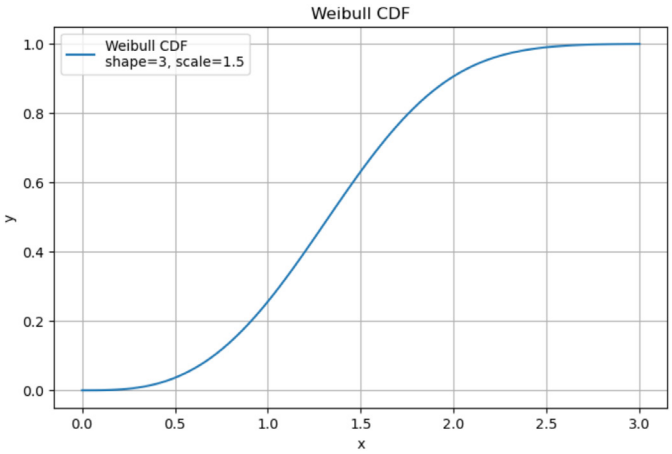


Figure 11: Weibull Comulative Density Function (author, 2024).

2.3.9 Batery Modeling

Battery storage systems operate by harnessing chemical energy and converting it into electrical energy, which is then stored for future use. The fundamental measure of a battery's storage capability is its energy capacity, typically quantified in kilowatt-hours (kWh). This capacity delineates the total amount of energy a battery can hold, with the usable capacity being the portion available for discharge, factoring in considerations such as depth of discharge (DoD). DoD is critical as it influences battery longevity; for instance, lithium-ion batteries maintain health when discharged up to 80-90% of their nominal capacity.

The power rating of a battery, expressed in kilowatts (kW) determines the maximum rate at which the battery can discharge energy. This rate impacts the battery's operational efficiency in various applications. The operation of these systems also includes the management of the state of charge (SOC), which is essential for maintaining battery health and efficiency. SOC limits, both upper and lower, are set to avoid excessive depletion, which could otherwise accelerate the degradation of the battery.

Round-trip efficiency represents the percentage of energy that can be retrieved from the battery as compared to what was stored, after accounting for losses due to self-discharge and other inefficiencies.

Not unlike photovoltaics, there is a plethora of methods to simulate and approximate the complex chemical processes of battery storage, with numerous simulation software options available. These methods range from intricate open-circuit voltage models, which consider the nonlinear relationship between the open-circuit voltage and the battery state of charge (SOC), such as the Shepherd model, which describes the voltage-current relationship with consideration of SOC. While these models are detailed and accurate, they require comprehensive datasheets of the analysed modules.

On the other hand, there are simplified linear approximations that provide a more generalized but less accurate representation of battery behaviour. These approaches are often preferred in scenarios where computational resources are limited or where a high degree of precision is not critical. However, the trade-off is a reduced ability to capture the nuanced dynamics of battery performance under varied operational conditions.

2.3.10 Batery Operation

The operation management of solar-coupled home battery systems is crucial in optimizing the balance between energy consumption, cost-efficiency, and grid stability. This management is predominantly achieved through algorithms that regulate when and how batteries are charged from solar panels and discharged to meet household energy needs or feed excess power back to the grid (von Appen et al., 2015).

The primary goal of operation management in these systems is to increase the efficiency of energy use while minimizing costs and mitigating the environmental footprint. Operation strategies aim to optimize the charging and discharging cycles of the battery based on various factors, including energy demand, solar energy production, electricity prices, and battery health. This involves making real-time decisions about when to store energy, when to use it directly, and when to sell it back to the grid (Angenendt et al., 2018).

Control strategies within these systems can be categorized into rule-based and predictive strategies (ibid.). Rule-based strategies typically operate under predefined rules that respond to changes in PV output, load demands, and grid electricity prices. For example, batteries might be charged when there is excess solar power and discharged when it is needed most, thereby increasing the self-consumption of generated solar energy. This strategy also allows for power to be fed back into the grid when there is a surplus, optimizing the financial returns through net metering or feed-in tariffs (Khezri, Mahmoudi and Haque, 2020).

Another sophisticated strategy employed is the dynamic price load shifting, where battery operation is optimized against electricity price fluctuations. Batteries are charged when electricity prices are low and discharged when prices are high, leveraging periods of differential pricing to reduce energy costs. This strategy requires real-time data on electricity prices and can significantly enhance system profitability and efficiency (Zhang et al., 2017).

A more advanced and increasingly common method involves Model Predictive Control (MPC). MPC uses forecasts of solar production and household energy demand to optimize battery charging and discharging over a future time horizon. This approach aims to align battery operation with predicted energy flows, thus ensuring that storage is used efficiently to cover energy needs with minimal reliance on the grid. MPC can be tailored to focus on various objectives, such as minimizing energy costs, balancing production and consumption, or reducing grid dependency (Angenendt et al., 2018).

Within these frameworks, various machine learning techniques enhance the predictive accuracy and operational efficiency of the battery systems (Guan et al., 2015). Reinforcement learning can adaptively tune the control strategies based on observed system performance and change environmental conditions. It does not rely strictly on precise long-term forecasts but can operate under uncertainty, making them particularly robust for residential energy management where solar production and household consumption can be unpredictable (ibid.).

Moreover, deep learning models, such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory

networks (LSTMs), are employed to forecast load and solar production more accurately. These forecasts feed into the MPC and other predictive strategies, enhancing the responsiveness and effectiveness of the operational management system (Corte Real et al., 2024).

2.3.11 Batery Degradation

Lithium-ion batteries are subject to degradation mechanisms that reduce their performance and lifespan over time. This degradation is predominantly caused by complex chemical and physical processes such as electrode material fatigue, loss of active material or electrolyte decomposition. They are influenced by various operational conditions including temperature, state of charge, charge/discharge rates, and depth of discharge. The two main degradation modes are cyclic aging pertaining to the wear and tear that batteries undergo due to charging and discharging cycles and calendar aging refering to the degradation that occurs as a function of time driven by chemical reactions within the battery.

Narayan et al., (2018) has distinguashied two categories, of battery lifetime estimation: performance-based and cycle counting models.

Performance-based models can utilize experimental techniques that monitor and quantify battery degradation over time, or they may employ semi-empirical methods that describe fading mechanisms through equations tailored to specific cell types. The limitation of these approaches is that they are designed for particular cells, within defined environmental conditions, and are validated over a constrained time period.

Cycle counting models differ from performance-based models by calculating parameters related to their end-of-life, including ampere-hour throughput, cycle count, or age since manufacture. These models are largely dependent on data from manufacturers and assume consistent ampere-hour throughout the battery's life, influenced by specific stressors such as depth of discharge and temperature.

In order to estimate the degradation based on the cycling of the battery charge a method to count the charge-discharge cycles is needed. The main method described in the literautre is the rainflow algorythm (Zhang et al., 2017), a technique primarily used in structural engineering for fatigue analysis. The algorithm starts by identifying all the local maxima and minima in the SOC full loading and unloading sequence. These points represent the turning points in the battery's charge and discharge process. Then the algorythm defines full cycles by matching each peak with a corresponding minimum. Additionally it forms half-cycles by pairing each point in the sequence with subsequent minima and maxima. The extracted cycles are then analyzed to count how many complete and half cycles occur. Each complete cycle typically consists of a charge and a discharge that returns to or below the starting SOC level. This has been visualised below.

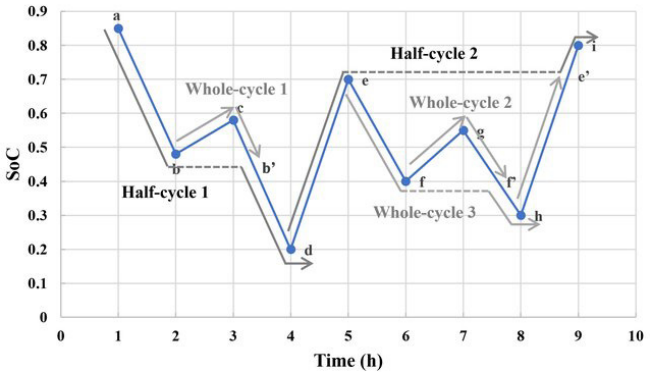


Figure 12: Rainflow Counting of Battery SOC cycles (Luo et al., 2021)

2.3.11 Conclusion

This subchapter outlines the critical role of solar radiation estimates and the influence of cell temperature on PV efficiency, showcasing the necessity of considering environmental factors such as temperature variations and wind speed in system design. It underscores the diversity of modelling techniques available, each with its advantages and limitations in terms of complexity and accuracy. While single-diode and multi-diode models provide a more granular perspective on the electrical characteristics of PV cells under varying conditions, simplified methods offer broader, albeit less precise, insights for quick assessments. Various types of losses are also highlighted. The objective of this project is to establish a methodology for informed financial decision-making, which necessitates a thorough approach to PV system modelling. Hence, the single diode model will be employed, incorporating all of the aforementioned loss factors.

The models for PV degradation and service life exhibit varying levels of implementation complexity and consider different factors that influence performance. Given the context of this project, which focuses on employing reinforcement learning for decision-making regarding the installation of PV panels, it is prudent to incorporate these models within the simulation environment. This approach will expose the decision-making agent to a spectrum of scenarios and probabilities reflecting real-

world variables. Therefore, degradation will be modeled stochastically, and panel longevity—or survival rate—will be quantified using the Weibull distribution. This method ensures that the model adequately reflects the uncertainties associated with the performance and durability of PV systems, enhancing the robustness and reliability of the decision-making process.

Furthermore, the battery system modeling and operation models are described comprehensively. These foundational models set the stage for the detailed battery modeling and operational strategies that will be elaborated in Chapter 6. Given the time constraints of this project, a simpler linear approximation has been chosen to model battery functionality. This method provides a straightforward yet effective way of estimating battery performance over time. Similarly to the approach used for modeling the degradation of PV modules, a stochastic process will be employed to simulate both the degradation and failure mechanisms of the battery. This process will utilize cycle counting as a key variable, allowing for the probabilistic modeling of battery wear and tear based on the number of charge-discharge cycles completed.

2.4 Reinforcement Learning

Reinforcement Learning is a paradigm of machine learning which emphasizes the learning of optimal behaviour through interactions with an environment. In this framework, an agent learns to make decisions by performing actions and observing the outcomes in terms of rewards or penalties. Unlike supervised and unsupervised learning, which rely on a pre-defined dataset, RL requires the agent to discover through trial-and-error which actions yield the most reward by exploring and exploiting the state space. Reinforcement learning is specialized in solving multi-stage decision problems (Sutton & Barto, 2020).

2.4.1 Markov Decision Process

The mathematical framework for modelling RL environment is usually formulated by Markov Decision Processes (MDPs). In this interactive learning process, at the discrete time *t(t=0,1,2,3…)* the agent observes the state $s_t \in S$, and takes an action $a_t \in A$, receives a numerical reward $r_t + 1 \in R$, and the state $s_t \in S$ moves to a new state, $s_t + 1 \in S$ (fig. x).

A Markov Decision Process is typically characterized by the five-tuple (S, A, P, R, γ), where P represents the state-transition probability function $P(s'|s,a)$ (with $s', s \in S$ and $a \in A$) depicting the state evolution's uncertainty based on the agent's actions. The discount factor γ in the range [0, 1] weighs the significance of immediate rewards against long-term rewards (Fu et al., 2022).

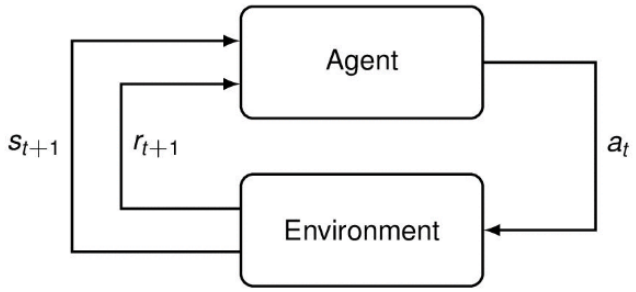


Figure 13: Markov Decision Process (Sutton & Burto, 2020)

2.4.1 Policy and Value

In RL an agent learns its optimal policy π through interactions with the environment. A policy π essentially maps states to the probabilities of choosing each available action. Specifically, at each time step t, the agent observes a state, executes an action, and receives a reward and transitions to a new state. These observations and rewards are utilized to refine the policy. Until the policy converges, the aforementioned procedure is repeated. This often involves balancing exploration (trying new actions to discover their effects) and exploitation (using known information to maximize rewards). In its simplest form this can be expressed as:

$$\pi(s, a) = \Pr(a = a \mid s = s) \qquad \textit{eq. 14}$$

Focusing only on the immediate reward signal results in always choosing the action that offers the highest immediate reward. To consider both immediate and future rewards (the sum of rewards from the present to the goal), a value function is created to represent the overall value of these rewards This is described by the state-value function for policy π:

$$V_\pi(s) = \mathbb{E}(\textstyle\sum_t \gamma_t r_t \mid s_0 = s), \qquad \textit{eq. 15}$$

where:
$V_\pi(s)$ represents the total accumulated reward when starting in state s and following policy π
$\sum_t \gamma_t r_t$ is the sum of discounted rewards over time,
$s_0 = s$ specifies a condition that the starting state is s.

In other words, the value of a state s, given the policy π is the expectation of how much reward will be obtained in the future given the current state and policy (Brunton & Kutz, 2022).

2.4.2 Classification of RL Algorithms

The landscape of RL algorithms is remarkably diverse, reflecting the wide array of approaches and techniques developed to address the challenges inherent in learning from interaction with an environment. This diversity stems from different algorithmic foundations, objectives, and application areas. A non-comprehensive classification with examples is given by figure 14.

2.4.3 Model-based vs Model-free

In model-based RL, the agent has access to or learns a model of the environment. This model predicts how the environment will respond to different actions, providing information about state transitions and rewards. The agent uses the model to "look ahead" and evaluate the consequences of actions before they are taken. This can be done using algorithms like Dynamic Programming, where the value Iteration updates the value function using the Bellman equation or Model Predictive Control.

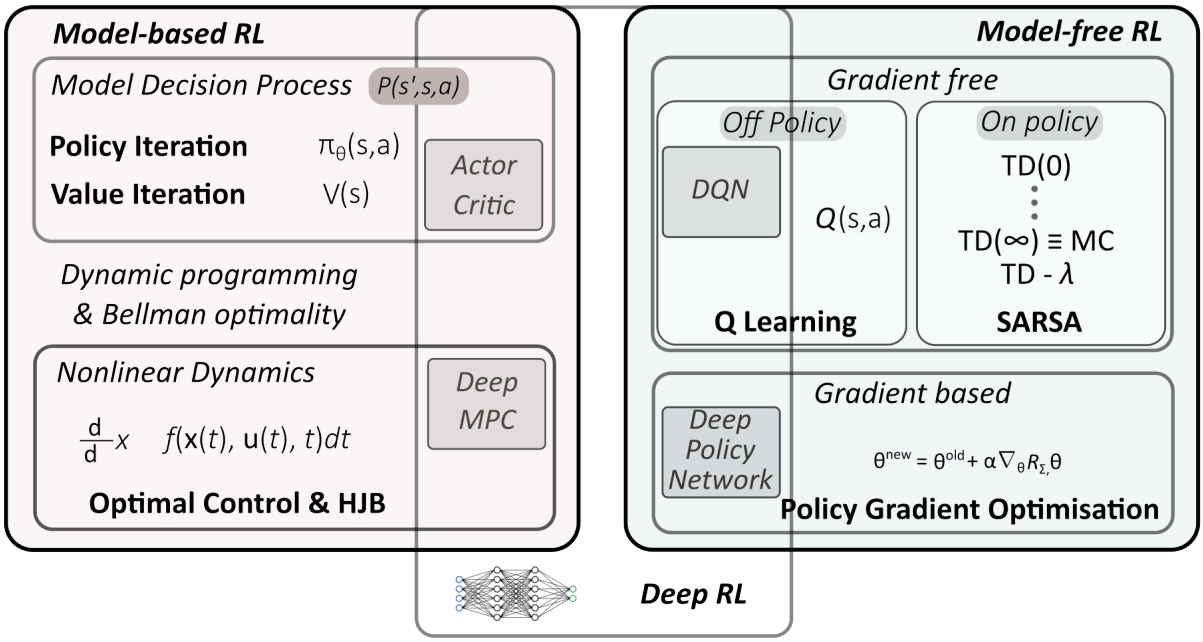


Figure 14: Reinforcement Learning Classification (author, 2024)

Model-free methods, on the other hand do not need to know environment models (i.e., state-transition probability function $P(s'|s, a)$ or $P(s'r\,|\,s, a)$ since they learn policies based on the information directly interacted with unknown environments.

2.4.3 Value–based vs Policy-based Algorithms

Value-based methods in model-free RL focus on estimating the value of states or state-action pairs rather than directly optimizing the policy. These methods are crucial in scenarios where it is more efficient or practical to learn the value of actions and derive the policy from these values. One of the essential value-based reinforcement learning approaches are tabular methods, where agent's policy is guided by a Q-table, where each action possesses a corresponding value. In the table, state and action are used as two indicators, and the action in each state corresponds to a Q-value. In Q-learning the Q-values are updated using the formula:

$$Q(s,a) \leftarrow Q(s,a) + \underbrace{\alpha[r + \gamma \max_{a'} Q(s',a') - Q(s,a)]}_{\textit{Temporal Difference (TD) Error}} \quad \textit{eq. 16}$$

where:
α is the learning rate,
r represents the immediate reward,
γ is the discount factor.

Another tabular-based method is SARSA (State Action Reward State Action), defined by:

$$Q(s,a) \leftarrow (s,a) + \alpha[r + \gamma Q(s',a') - Q(s,a)] \quad \textit{eq. 17}$$

The term $Q(s',a')$ involves the action a' that the current policy dictates in the new state. SARSA is an on-policy algorithm. It learns the value of the policy being followed, including the exploration steps. On the other hand, Q-learning is off-policy. It updates its estimates based on the assumption that the best action is chosen, regardless of the action actually taken.

Unlike value-based methods, which first estimate the value of states or actions and then derive a policy, policy-based methods optimize the policy function directly. Policy-based methods update its function approximator according to the gradient of expected reward, with respect to the policy parameters θ. The following equation represents a general update rule in policy-based RL:

$$\theta_{new} = \theta_{old} + \alpha \nabla_\theta R(\Sigma, \theta) \quad \textit{eq. 18}$$

where:
α is the learning rate, determining the size of the update step,
∇_θ is the gradient with respect to the policy parameters θ,
$R(\Sigma,\theta)$ represents the expected reward.

Some algorithms as the Actor-Critic methods combine both value and policy-based methods. There, the 'actor' updates the policy distribution in the direction suggested by the 'critic,' which estimates the value function. Thus reducing the variance in the policy updates and ensuring a continuous policy improvement.

2.4.4 Deep Reinforcement Learning

Figure 12 illustrates that various RL algorithms can be classified as "deep," independent of their specific type. Essentially, an RL algorithm is deemed deep if it employs artificial neural networks (ANNs) in its policy or value functions. In recent years Deep RL (DRL) has gained significant attention due to its ability to tackle complex problems with high-dimensional input spaces (Brunton & Kutz, 2022).

An ANN is a computational model inspired by functional aspects of biological neural networks. It consists of interconnected nodes, or "neurons," which are organized in layers. These layers include an input layer, one or more hidden layers, and an output layer. Each neuron within these layers is linked with weighted connections that are adjusted during the learning process. ANNs are designed to recognize patterns in data through a process of learning from examples, making them particularly useful in tasks such as classification, regression, and pattern recognition. By adjusting weights based on the input data, ANNs can improve their performance and make predictions or decisions without being explicitly programmed for the task (Chen et al., 2023).

In fact, all the aforementioned RL algorithms have their "deep" counterparts, such as:

Deep Q-Networks (DQN) using an ANN to approximate the Q-value function.

Proximal Policy Optimization (PPO) where a ANN is used to directly parameterize the policy, outputting the probability distribution of actions.

Advantage Actor Critic (A2C) employs two neural networks - one for the Actor (policy) and another for the Critic (value function).

Deep Deterministic Policy Gradient (DDPG) combining ideas from DQN and policy gradients, using a neural network to represent a deterministic policy and another network for the value function.

2.4.5 Single agent vs. Multi-agent RL (MARL)

Not shown in the diagram, and more pertinent to the application of the algorithm rather than its mathematical framework, is the distinction of whether the RL model is pursuing one or multiple reward signals. It is important to notice that single agent algorithms can also optimise for multiple objectives, however that requires designing a synthetic reward function as a weighted sum of different objectives.

Multi-agent RL involves multiple agents simultaneously interacting within the same environment. These agents can be cooperative, competitive, or a mix of both. They are inherently more complex than single-agent RL. In single-agent RL, since policies must adapt dynamically to the strategies of other agents. All aforementioned algorithms are single-agent. MARLs include: Multi-Agent Deep Deterministic Policy Gradient (MADDPG) or Multi-Agent Proximal Policy Optimization (MAPPO) (Hayes et al., 2022).

2.4.6 RL for Building Energy-use Related Decision-making

In current research, DRL is being utilized to enhance the efficiency and decision-making processes in building energy systems. These applications range from optimizing operational costs and energy consumption in individual building energy subsystems to aiding in the adoption of renewable energy at various scales, including residential, municipal, urban, and regional. Table 4 presents a summary of the RL algorithms that have been employed in these areas.

RL is applied to a variety of energy-related domains such as smart homes, school buildings, commercial buildings, and photovoltaic systems. In general, it can provide customized solutions for different energy systems, taking into account the unique characteristics of each system's operation and the specific objectives set by the operators or owners. The applications are range from improving energy efficiency (Stripp, 2023), reducing consumption (Shuvo & Yilmaz, 2022) to optimizing for various objectives such as user comfort (Cheng et al., 2016), financial gain (Jung et al. 2021) or reducing human effort.

The single-objective problems typically focus on reducing energy consumption in a variety of settings, such as residential buildings, school buildings, and commercial structures. Some specific applications include improving the energy efficiency of smart home appliances (Xu et al.

RL Algorythm	Category	Problem	Application	Reference
DQN (Deep Q-Network)	Model-free, Value-based, off-policy, single-agent	Single-objective, Discrete action space	Reducing energy consumption in commercial buildings with human-in-the-loop.	Wei et al. (2020)
		Single-objective, Discrete action space	Reducing the energy consumption of a school building with natural ventilation.	Stripp (2023)
		Single-objective, Discrete action space	Maximising financial gains of a PV system.	Jung et al. (2021)
		Single-objective, Discrete action space	Reducing indoor pollution with natural ventilation.	An et al., (2021)
Multi-DQN (Multi-Deep-Q-Network)	Model-free, Value-based, off-policy, multi-agent	Multi-objective, Discrete action space	Minimising the necessary time of EV charging at a residential charging station.	Zhang et al. (2020)
PPO (Proximal Policy Optimisation)	Model-free, Policy-based, on-policy, single-agent	Single-objective, Discrete action space	Maximising financial gains of a PV system.	Jung et al. (2021)
		Single-objective, Continuous action space	Capacity Scheduling for PV-Battery Storage System	Huang & Wang, (2021)
A3C (Asynchronous Advantage Actor-Critic)	Model-free, policy-based and value-based, single-agent	Single-objective, Continuous action space	Reducing the energy consumption of through smart building control algorithm.	Zhang et al. (2021)
EB-C-MADRL (Entropy-Based Collective Multiagent DRL)	Model-free, Value-based, off-policy, multi-agent	Single-objective, Continuous action space	Reducing the energy consumption of a smart home in a microgrid.	Yang et al. (2019)
DDQN (Double Deep Q Learning	Model-free, Policy-based, off-policy, single-agent	Single-objective, Continuous action space	Reducing the energy consumption of hydronic heat pump	Zheng (2022)
DDPG (Deep Deterministic Policy Gradient)	Model-free, Policy-based, off-policy, single-agent	Single-objective, Continuous action space	Reducing the energy consumption of hydronic heat pump	Zheng (2022)

Table 04: Overview of RL applications for Building Energy Managment Related Purposes

RL Algorythm	Category	Problem	Application	Reference
DDPG (Deep Deterministic Policy Gradient)	Model-free, Policy-based, off-policy, single-agent	Multi-objective, Continuous action space	Minimisation of HVAC energy consumption and maximising thermal comfort in an office building.	Gao et al. (2020)
SAC (Soft Actor-Critic)	Model-free, Policy-based, off-policy, single-agent	Multi-objective, Continuous action space	Reducing the energy consumption and maximising user comfort in a in solar microgrid using battery energy storage systems	Onile et al. (2022)
Multi-Agent Q-Learning	Model-free, Value-based, off-policy, multi-agent	Single-objective, Discrete action space	Reducing the energy consumption of smart home appliances including real data of electricity price and PV generation	Xu et al. (2020)
MDP	Model-based	Single-objective, Continuous action space	Maximising visual comfort in an office by optimising lighting Conditions.	Park et al. (2019)
Q-learning	Model-free, Value-based, off-policy, multi-agent	Multi-objective, Discrete action space	Maximum power point tracking for photovoltaic power generation.	Cheng et al. (2016)
	Model-free, Value-based, off-policy, single-agent	Multi-objective, Discrete action space	Maximum power point tracking for photovoltaic power generation	Zhang et al. (2019)
FQI (Fitted Q-iteration)	Model-free, Value-based, off-policy, single-agent	Single-objective, continuous action space	Reducing the maximum power injection from PV systems to the grid from HVAC use.	Mbuwir et al. (2017)

natural ventilation systems (An et al., 2021), photovoltaic power generation (Mbuwir et al., 2017), and HVAC systems (Gao et al., 2020).

The multi-objective problems involve not only reducing energy use but also maximizing factors like user comfort or financial gains. For instance, Multi-DQN has been used to minimize the necessary time of electric vehicle charging at residential stations (Zhang et al., 2020), while algorithms like SAC (despite being a primarily single-agent algorithm) have been applied to optimize comfort in conjunction with energy savings in solar microgrid systems (Onile et al., 2022).

The model-free, value-based, and off-policy algorithms like DQN, Multi-DQN, DDQN, and Multi-Agent Q-Learning are predominantly used for discrete action spaces in single or multi-agent settings. These algorithms have been effectively applied to optimize energy consumption in various building types and systems, including commercial and school buildings (Stripp, 2023), and photovoltaic PV systems (Jung et al., 2021).

On the other hand, model-free, policy-based, off-policy algorithms like PPO, A3C, DDPG, and SAC, as well as the model-based HERS and the model-free FQI, have been employed for both discrete and continuous action spaces. These are particularly useful in complex scenarios as maximizing comfort and energy efficiency in smart homes (Huang & Wang, 2021) and microgrids (Yang et al., 2019).

2.4.7 Continuous vs. Discrete Action Space

In general, the literature indicates a distinction between the suitability of different algorithms to discrete and continuous actions spaces. Discrete action spaces consist are these consisting of a finite set of actions ex. windows closed or opened or a PV panel installed or not. While continuous action spaces present an infinite number of possible actions ex. optimising the rate at which fresh air is introduced into the building.

The discrete nature facilitates a straightforward selection of actions by directly comparing the estimated values for each possible action and choosing the one with the maximum value (Brunton & Kutz, 2022). Q-learning and its variants especially thrive in discrete environments by approximating the optimal action-value function. Other methods for instance (as PPO) can be used in for both discrete and continuous tasks. Figure 15 provides an overview of RL algorithms and their suitability.

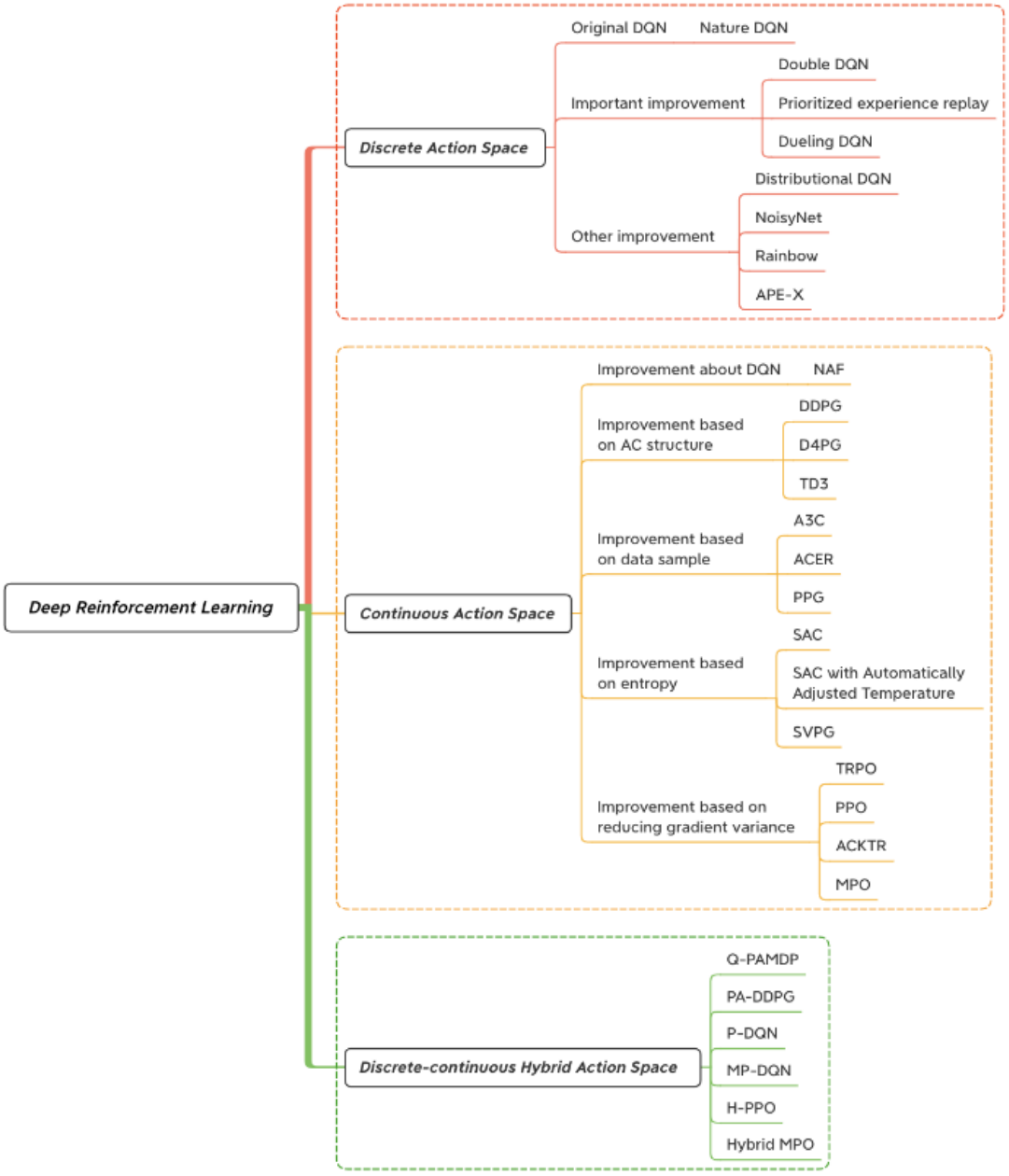


Figure 15: Suitability of RL Algorythms for Disrcete and Continous Action Spaces (Zhu et al., 2022)

2.4 Variables Modelling

When building a reinforcement learning model, it is vital to simulate a range of future variables scenarios for training. Such diverse exposure enhances the model's ability to manage real-world uncertainties, improves its generalisation ability, helps to prevent from overfitting and improves decision-making by identifying patterns in the variables and optimal times for PV panel implementation and replacement (Sutton & Barto, 2020). Electricity prices, CAPEX and the electricity grid emissions tend to evolve unpredictably given the inherent volatility influenced by market trends, policy changes, and environmental factors. Precise long-term energy price or PV module price predictions are challenging; this thesis aims to realistically simulate long-term price fluctuations rather than achieve exact precision.

2.4.1 Conumer Energy Prices

Weron (2014) conducted an extensive review of electricity price forecasting methods, distinguishing various approaches through a detailed classification. The methodologies of these approaches have been briefly summarized below:

Fundamental Methods

The dynamics of electricity pricing are modeled by simulating the impact of critical physical and economic factors. This involves predicting key influencers like demand, weather, and system specifics. Applying these models is complex due to data limitations and the challenges of factoring in the random variations of main drivers.

Multi-agent Models

They mimic the dynamics of a system with various agents (like power units and firms), simulating price formation by aligning market demand and supply. These models face difficulties in validating electricity markets' assumptions and accurately pinpointing key participants and their tactics.

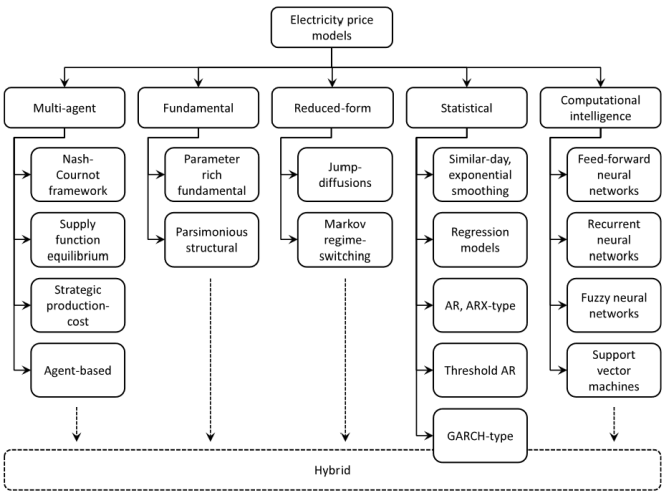


Figure 16: An Overview of Electricity Tariff Prediction (Weron, 2014)

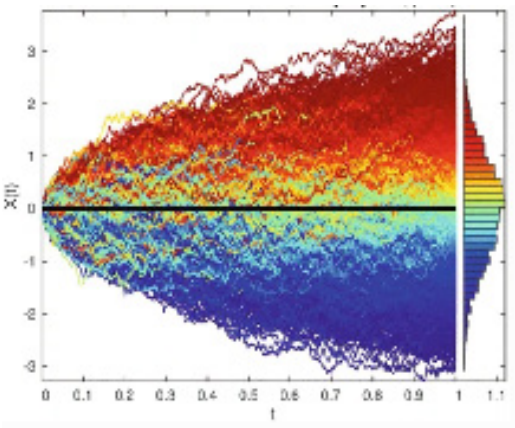


Figure 17: Geometric Brownian Motion in action (Petters and Dong, 2016)

Reduced-Form Methods

These models analyze the time-based statistical traits of electricity prices, focusing on evaluating derivatives and risk management. Rather than providing exact hourly forecasts, they aim to replicate essential attributes of daily electricity prices, like future marginal distributions. Their main strengths are their simplicity and the ease of tracking analysis results.

Statistical Methods

These methods employ statistical techniques or modified econometric models to forecast power market trends, combining historical prices with current or past external data like consumption, production, or weather. Their precision hinges on algorithm performance and data integrity.

Computational Intelligence

These methods integrate learning, evolutionary, and uncertain elements to craft adaptable strategies for complex systems. Examples are ANN, support vector machines, and genetic algorithms. Their complexity and the need for vast, high-quality data make them unsuitable for this thesis.

2.4.2. Stochastic Processes Simulation

From this overview it becomes clear that reduced – form, models, which can be calibrated to historical data are the most suitable for the problem in question. This approach simplifies the complex mechanisms behind energy pricing into a more manageable form without losing the essence of market dynamics (Weron, 2014).

Under this framework the increments of the electricity tariff, denoted as X_t can be derived from specific instances of a broader stochastic differential equation (SDE)(Ibe, 2013):

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t + dq(X_t, t) \qquad \textit{eq. 18}$$

where:
dX_t is the incremental change in the spot electricity price at time t,
$\mu(X_t t)dt$ is a function representing the drift rate of the price. It models the predictable, trend following part of the price change,
$\sigma(X_t t)dB_t$ is a function representing the volatility of the price with a Brownian motion used to represent random motion,
$dq(X_t t)$ is a jump process, capturing sudden and significant changes in the price.

SDE incorporates both a deterministic component (like in ordinary differential equations) and a stochastic component to model the randomness. Different variants of this equations have been found in literature to generate future electricity tariffs based on data obtained from the past.

Geometric Brownian Motion (GBM) is a stochastic process commonly used in financial mathematics (Farida Agustini et al., 2018) and to simulate future energy pricing scenarios (Penizzotto et al., 2019). This is due to its ability to incorporate fixed rates of drift and volatility, reflecting common tariff patterns. In GBM, these coefficients are typically constants or proportional to the current price, without any jump component. GBM values exhibit positive skewness and follow a log-normal distribution, with variance that grows over time (Weron, 2014).

The price change process in GBM is a mix of a deterministic drift effect and a stochastic volatility effect. Tariff denoted as X_t starts with an initial value $X_0 > 0$ and follows a specific SDE:

$$dX_t = \mu X_t dt + \sigma X_t dB_t \qquad \textit{eq. 19}$$

Kaminski (1997) employs **Merton's jump-diffusion model**, which merges GBM with a jump process. It integrates the continuous path of GBM with sudden, discrete changes, representing more abrupt shifts in the underlying variable.

$$dX_t = (\mu - \lambda k)X_t dt + \sigma X_t dB_t + dJ_t \qquad \textit{eq. 20}$$

where:
dJ_t represents the jump component at time t, and is an increment of a homogenous Poisson process.

Its primary limitation is its oversight of mean reversion or the tendency to return to a 'normal' price level. Due to the jump component, in instances of a price spike it perceives this new price as standard and continues to evolve randomly (Weron, 2014).

More advanced models using the Merton's jump-diffusion model have been developed by Albanese et al. (2012) and Cartea and Figueroa (2005), where a time-dependent volatility σ have been taken into account. While Weron (2008) and Bhar et al. (2013) have calculated the jump component with a periodic intensity function to account for the seasonality of the tariff spikes.

2.4.3. Scenario Generation for the Electricity Tariff

In order to generate the multiple scenarios for the RL model training Jung et al. (2021) creates multiple Monte Carlo simulations using GBM relying on the stochasticity of the Brownian Motion. Here the volatility was calculated using the historical consumer price data and the mean drift is derived based on the assessed volatility.

Similarily Penizzotto et al., (2019) have used GBM to simulate paths of the electricity tariff over the period of 35 years, arguing that setting the drift and volatility rates for standard tariff behaviors is a straightforward process.

In general GBM can be approximated for a discrete timestep under the analytic solution to its SDE for an arbitrary initial value X_t (under Ito Interpretation) described as (Ibe, 2013), assuming that the parameters are constant:

$$X_t = X_0 exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right) \qquad \textit{eq. 21}$$

where:
X_0 is the price of electricity at the start of the process,
μ is a constant drift rate coefficient,
σ is a volatility coefficient of the process,
B_t is a Brownian motion.

Having obtained the past data of consumer energy prices the μ and σ can be estimated by calculating log returns under the assumption that they are normally distributed (Stander, 2020), with mean and variance equal to μ and σ^2:

$$E[\log(X_t) - \log(X_{t-1}) = \mu \qquad \textit{eq. 22}$$

$$var[\log(X_t) - \log(X_{t-1})] = \sigma^2 \quad \textit{eq. 23}$$

$$\log(X_t) - \log(X_{t-1}) \sim N(\mu, \sigma^2) \quad \textit{eq. 24}$$

2.4.4 Net-Metering

The net-metering rate is directly tide to the electricity tariff. However in the Netherlands starting from 2025, net-metering is going to be phased out under the amendment to the Sustainable Energy Production Incentive Scheme ("Stimuleringsregeling Duurzame Energieproductie") until 2031 (Ministerie van Algemene Zaken, 2023) (fig. 18).

Nevertheless, energy suppliers will provide reduced compensation. Government proposals specify a reasonable compensation ("redelijke vergoeding") as the minimum rate, approximately 30% of the base electricity price, excluding taxes and surcharges, as assumed in this analysis (van den Berg, 2023).

Additionally, it is important to acknowledge that, as of January 2024, certain energy suppliers have implemented an additional fee ranging from 10 to 30 euros per month for customers with solar panels. This surcharge is intended to compensate for substantial imbalance costs incurred when an unforeseen quantity of energy is introduced into the grid (Leone, 2023).

2.4.5 PV CAPEX

Literature on the stochastic scenario generation of photovoltaic panel cost is considerably less abundant than that on electricity tariffs.

The literature review reveals four pertinent methods. Two of them utilise the concept of learning curves, which are commonly used to forecast future prices of technologies, by showing how investment costs per unit of capacity (Wp in case of PV panels) decrease as the total installed capacity grows. A key concept in these curves is the learning rate, (LR) which represents the rate at which costs fall when cumulative capacity doubles (Kim et al., 2019). According to Swanson's law price of solar photovoltaic modules tends to drop by 20% for every doubling of cumulative shipped volume (Swanson, 2006). The long-term LRs of the photovoltaic and BESS technologies for Europe has been described by Vartiainen et al., (2019).

Ornstein-Uhlenbeck Process

Kim et al., (2019) have found that historical price data for PV modules appears to exhibit a mean-reversion characteristic, which is a propensity for the trend to gravitate back towards the average after jumps. The simplest model with this property is the Ornstein-Uhlenbeck (OU) Process given by the following SDE:

$$dX_t = \mu(\theta - X_t)dt + \sigma dW_t \quad \textit{eq. 25}$$

where:
X_t denotes the EG at time t,
θ is a long-term mean coefficient,
μ is a mean reversion rate,
σ is the volatility of the process,
dWt is a Brownian motion.

The authors do not explicitly use OU to generate future PV prices. However, they modify the OU process to incorporate the features of the learning curve for PV in order to model propability discribution of the future prices, which are then utilised for scenario trees.

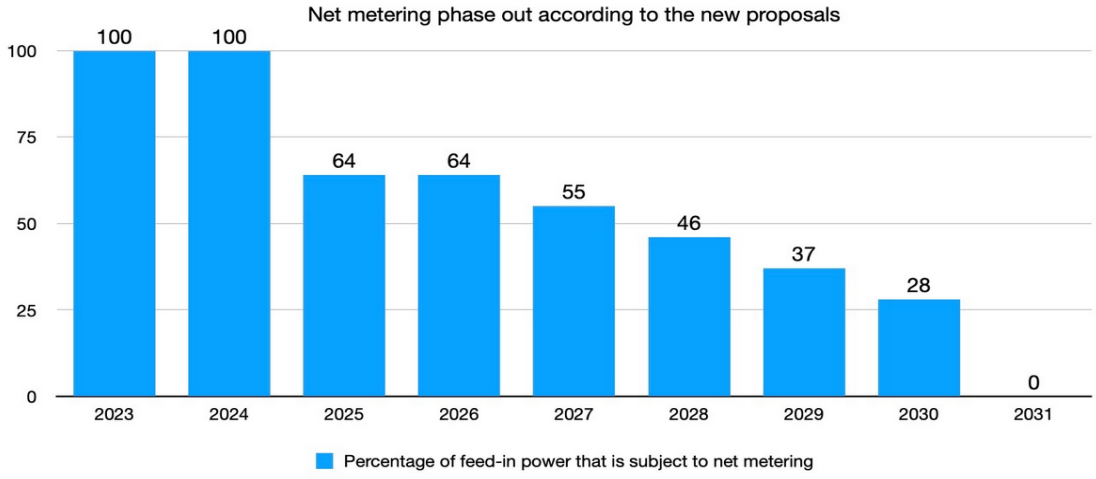


Figure 18: Net-metering rates in the Netherlands in the following decade (van den Berg, 2023)

GBM

A standard geometric brownian motion was used by Moon and Baran (2018). It was assumed that $\mu < r$ for convergance, where r represents the discount factor adjusted for risk.

GBM with a Learning Rate

Jung et al., (2021) have simulated future PV module prices based on the predicted future cumulative production. The cost of the PV module when the cumulative installation capacity reaches Q is given by:

$$C_Q = C_0 Q^{-\beta} \qquad \textit{eq. 26}$$

where:
C_Q is the price of the PV module when the installation capacity of the PV module is reached at Q,
C_0 is the initial cost of the PV module,
Q is the cumulative installation capacit in MW,
β is the learning rate, indicating the percentage decrease in price with each doubling of capacity (Elshurafa et al., 2018).

As Q increases, according to Swanson's Law, the price C_Q should decrease. The predicted total amount of solar installation capacity can be utilized to determine the future cost of photovoltaic modules and then to calculate the mean drift, volatility is estimated from historical data (ibid., 2021).

Jump Diffusion Process (GBM with a Poisson Events Model)

Penizzotto et al., (2019) regards CAPEX as a exogenous random process, influenced by technological advancements, characterised by a mixed stochastic process. This consists of a continuous Geometric Brownian Motion to account for regular market price variations, alongside a Poisson process to model the impact of technological innovations, economic cycles etc. It follows the methodology first developed by Merton (1976) for stock price precitction. The stochastic dynamics of CAPEX at time t are described as:

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t + \sum_{j=1}^{N_t} Y_j\, dJ_t$$

eq. 27

where:
$\mu(t,X_t)$ is the drift term,
$\sigma(t,X_t)$ is the diffusion coefficient,
N_t is a Poisson process with a rate λ,
Y_j represents the size of the j-th jump, independent and identically distributed
dJ_t is the differential of the jump process, which increments by 1 at each time a jump occurs

2.4.6 Grid Emission Factor

No literature regarding stochastic modelling or scenario generation of EG has been found.

2.5.7 New PV Performance Improvement

No literature regarding stochastic modelling or scenario generation of the improving efficency of PV modules has been found.

2.5.7 Battery CAPEX

No literature regarding stochastic modelling or scenario generation of the BESS CAPEX has been found.

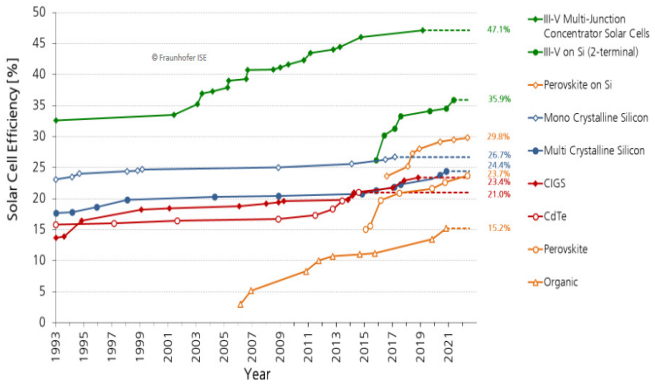


Figure 19: PV Efficency Improvement Over Time (Fraunhofer Institute, 2023)

2.5.8 Conclusion

It needs to be emphasised that despite the variety of available models, a singular approach may not be sufficient to address the intricacies of PV system cost and efficiency modeling. There is a need for a hybrid approach that combines the strengths of different models to achieve a more comprehensive and realistic training environment.

Despite no literature on stochastic modelling of both EG and PV csystaline cell improvement certain characteristic of those variables allow to pair them with different stochastic or propabilistic processes. This will be described in more detail in chapters 5 and 6.

2.5 Model Deployment

2.5.1 Reinforcement Learning Deployment

The primary objective of creating a machine learning model is to address a specific issue. Such a model can effectively achieve this only when it is deployed and utilized in a real-world setting. Model deployment refers to the act of transferring a machine learning model from a non-operational setting into a working production environment.

This involves two key tasks - creating the necessary infrastructure for operating the model, and developing the model in a way that allows for its use and maintenance (Afsar et al., 2022). There are to main deployment methods:

Online (real-time) - predictions or inferences occur instantly. The model analyses and interprets data as it is recieved.

Offline (batch) - predictions or inferences occur at regular intervals, on a batch of data accumulated since its last execution. Since it produces results with some delay, it is effective in scenarios where immediate or real-time model outcomes are not essential (Klushin, 2024).

In the realm of RL applied to decision-making, the implemented model becomes part of a recommendation system (RS). These systems consist of software designed to assist users in discovering items of interest or in making informed decisions. This is achieved by predicting their preferences, ratings of items, or estimating the outcomes of their choices (Afsar et al., 2022). The front end solution of a RS can be a web interface, mobile app or system software.

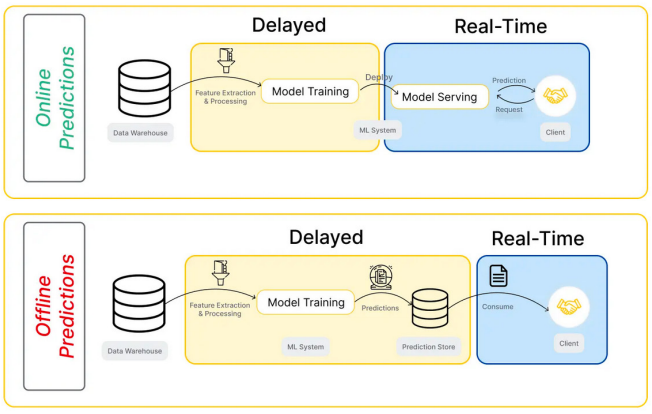


Figure 20: Offline and Online Deployment (Klushin, 2024)

2.5.2 Explainability

The RS gives periodical recomendetion to the user. It is important however that the user understands why a certain recomendation has been made. Explaining recommendations can enhance the user experience, increase their confidence in the system, and assist them in making more informed choices (Afsar et al., 2022). Chen et al., (2023) divides explainable methods in terms of application stage:

Ante-hoc - involve designing models that are inherently interpretable during their training process. This approach is difficult to implement in complex DRL models.

Post-hoc - are methods applied to complex, "black-box" models after they have been trained. Their purpose is to explain how certain input features affect the model's output.

The other distinction is about the interpretability scope:

Global Interpretation - involves comprehensively understanding how a model makes predictions, focusing on the overall impact of input features on predictions.

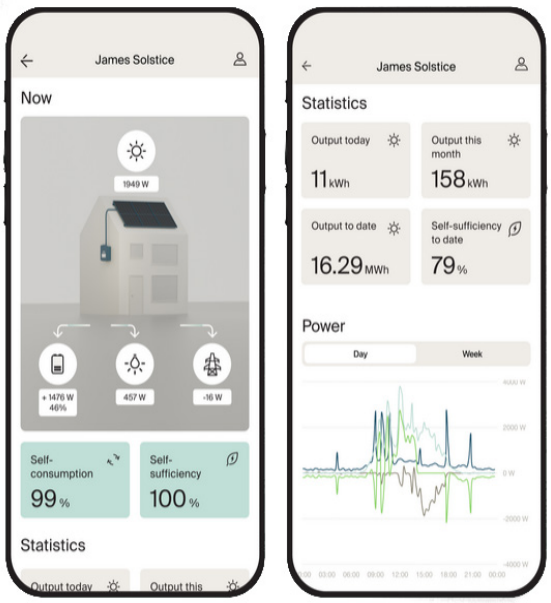


Figure 21: Autarco Monitoring (Autarco , n.d.)

Local Interpetation - offers a clear insight into how a model arrives at a prediction for a particular data instance.

It emphasizes the role of individual features in that specific prediction, rather than a broad overview of feature importance.

The Google PAIR team. (2019) have articulated a number main considerations in terms of model interpretability:

- explaining the datasources - prevents concerns about privacy and assists the user in understanding when to use their own judgement,
- linking explanations to user actions and the resulting system responses, along with clear explanations, can assist users in gradually building an appropriate level of trust
- offering thorough explanations and disclosing the reasoning for predictions made with different levels of confidence confidence.

Authors describe that appriopirate explanations should ideally be given in-the-moment, yet if information is not sufficient the user should be provided with external resources and the explenations could be contextualised. Additionally, it is stresed that another way of building a strong connection between RC and the end user is providing them with the ability of specifing their preferences regarding the model's output.

2.5.3 PV Monitoring Systems

PV monitoring systems, which are installed alongside the solar panels, provide detailed information about the panels' performance through software applications or web portals, by gathering data from the inverters of the solar array. They showcase panel's functioning to the owner and inform them about pottential faults. The equipment in these systems is designed to identify changes in weather conditions and sunlight levels (Samsukha, 2023).

Several companies, including SolarEdge, Enphase, and Autarco, have implemented monitoring systems for PV installations. These systems are designed to provide an attractive graphical representation of the PV system and offer users detailed information about both current and

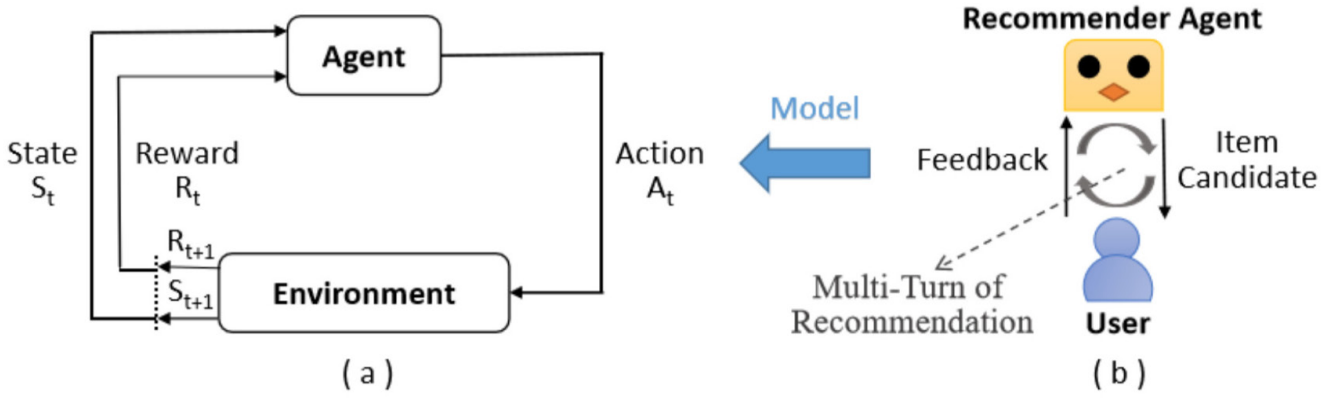


Figure 21 B) RL-based recommender system models the interactive reccomendation task as a MPD (Afsar, Crump and Far, 2022).

03. WORKFLOW PLAN

03 WORKFLOW

3.1 Selection of the DRL Algorythm

The RL algorithms found in the literature were thoroughly analysed for the suitability for the problem disclosed in this thesis considering the following criteria:

Convergence Speed i.e. how quickly the algorithm learns an effective policy.

Robustness and Stability i.e. its performance consistency across different scenarios and its resilience to changes in environmental conditions or market prices.

Scalability i.e. how well the algorithm handles increasing complexity, such as more panels or varying environmental conditions.

Multi-Objective Optimization Capability i.e. examine the algorithm's ability to handle multi-objective optimization effectively, ensuring that neither financial gain nor environmental impact is disproportionately prioritized.

Access to resources i.e availability and quality of learning resources, tutorials, and community support for the algorithm. This is crucial for understanding its nuances and troubleshooting issues.

The two selected algorithms are:

3.1.1 Proximal Policy Optimisation (PPO)

PPO (Proximal Policy Optimisation) is a model-free, policy-based, on-policy deep reinforcement learning algorithm. PPO leverages an actor-critic network, offering enhanced performance through rapid, data-responsive learning (Schulman et al., 2017). This network is particularly effective in handling large state dimensions, ensuring efficient convergence. This capability makes it well-suited for complex optimization problems like the one in this study, where a multitude of variables need to be taken into account (Jung et al., 2021).

The core of PPO is its objective function, which is designed to minimize large policy updates, preventing destructive steps that could harm learning progress. The clipped surrogate objective function $L^{CLIP}(\theta)$ (*eq. 29*) is a crucial part of this. It uses a clipping mechanism to limit the amount by which the new policy can diverge from the old policy. The probability ratio $r_t(\theta)$ (*eq. 30*) compares the probability of an action under the current policy π_θ to its probability under the previous policy $\pi_{\theta old}$. The advantage function A_t measures how much better an action is compared to the average action at a given state. Alongside the policy optimization, PPO also updates a value function, typically using a squared-error loss L^{VF}. The final loss function combines these elements, represented as (*eq. 31*). Policy updates in PPO are done using stochastic gradient ascent, optimizing the parameters of the policy network to maximize the final objective function (Brunton & Kutz, 2022).

$$L^{CLIP}(\theta) = \mathbb{E}[min(r_t(\theta)A_t), clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t] \quad \textit{eq. 29}$$

$$r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta old}(a_t|s_t)} \quad \textit{eq. 30}$$

$$L^{PPO}(\theta) = \hat{\mathbb{E}}\left[L^{CLIP}(\theta) - c_1 L^{VF} + c_2 S_{[\pi_\theta]}\right] \quad \textit{eq. 31}$$

3.1.2 Deep Q-Network (DQN)

DQN (Deep Q-Network) is likely the most popular DRL framework. Unlike a standard Q-learning uses a NN to approximate the Q-value instead of the Q-table, calculating all the state-action values for a single in a single pass through the NN. This is a clear advantage as the NN can extrapolate across the range of state-action pairs. Such generalization enables an agent employing DQN to effectively estimate the values of state-action pairs during inference that are either rare or have never been encountered before during training (Alexandros Iosifidis and Anastasios Tefas, 2022, pp.121–123).

The agent's experiences at each time step, denoted as *(s,a,r,s')* are stored in a replay buffer. Random batches from this buffer are used to train the network. During training to balance the exploration and exploitation the agent uses an ϵ-greedy policy, which selects a random action with probability ϵ and the action with the highest Q-value (as predicted by the network) with probability $1-\epsilon$. To refine the estimated Q-values, episodes are drawn from the replay memory and a loss function based on equation *eq. 16* is computed as (ibid.):

$$L_{s,a,r,s'} = [r + \gamma \max_{a'} Q(s', a') - Q(s, a)]^2 \quad \textit{eq. 32}$$

During training the gradient of the loss function is then backpropagated through the DQN, leading to adjustments in its parameters.

3.2 Workflow Methodology

The experimental workflow for this thesis is divided into four steps:

A. Simplified Toy Problem

This phase functions as a preliminary testing environment to evaluate the proposed methodology and provides an opportunity for the author to become acquainted with the selected libraries, debug the system, and identify critical areas for improvement.

The toy problem simplifies the scenario by assuming constant variables during optimization. The PV panels are modeled using the single-diode model, although degradation and failure mechanisms are not incorporated at this stage. The primary aim is to scrutinize and validate the framework for constructing a RL environment, assess the effectiveness of the selected algorithms, refine the reward function, and evaluate the modeling of the PV system. These aspects are comprehensively detailed in this section. Successful convergence of the algorithm signifies the project's readiness to progress to the next phase.

B. PV System Optimisation without BESS

In this phase, the methodology integrates the stochastic and temporal characteristics of relevant variables. The primary focus for optimization is the size of the rooftop PV array, while battery energy storage is not considered in this scenario.

The methodology from the previous section is tested and evaluated iteratively following the implementation of each variable along with its associated uncertainties. This iterative process may necessitate modifications in the approach, particularly concerning key components of reinforcement learning such as the state and action spaces or the selection of the agent.

The stochastic processes utilized to simulate these variables are detailed in Table 05. The rationale behind the choice of these processes is more thoroughly described in Chapter 5. Should an approach prove unsuitable—due to technical issues or a lack of data—an alternative method will be explored and potentially implemented.

Furthermore, the approach will be validated using real household load profiles typical for residences in the Netherlands and employing various types of commercially available PV panels, considering a realistic time-of-use electricity tariff rate. This validation aims to ensure that the model accurately reflects the performance and feasibility of PV installations under realistic operational conditions. Two phases of the optimisation will be performed: financial only and financial and environmental combined. The results will be discussed and analysed.

C. PV System and BESS Optimisation

In this segment of the study, the optimization variables encompass both the PV array and the capacity of the BESS. The method for defining the observation space, actions, and rewards has been adjusted to meet the demands of this new optimization framework. Specifically, two new stochastic variables have been introduced to the model: the CAPEX for battery storage and its degradation process.

Moreover, a substantial part of this section is dedicated to outlining a reinforcement learning-based approach for simulating the operation of the battery system.

D. Deployment

The final phase of the proposed workflow involves developing a front-end recommendation system for the model. This will serve as a demonstrative interface, which could be potentially integrated into an actual PV monitoring system.

Variable	Modelling Method
Electricity Tariff	Geometric Brownian Motion
Net-Metering/ Feed in Tariff	Deterministic
PV CAPEX	Geometric Brownian Motion with Poisson Jumps
PV Cell Improvement	Gamma Process
PV Failure	Weibull Distribution
PV Degradation	Gaussian Distribution
Grid Factor	Ornstein-Uhlenbeck Process
PV Module Embodied Carbon	Deterministic
Battery CAPEX	Geometric Brownian Motion with a Learning Curve
Battery Degradation	Non-homogenous Gamma Process

Table 05: Relevant Variables and their simulation methods

03 WORKFLOW

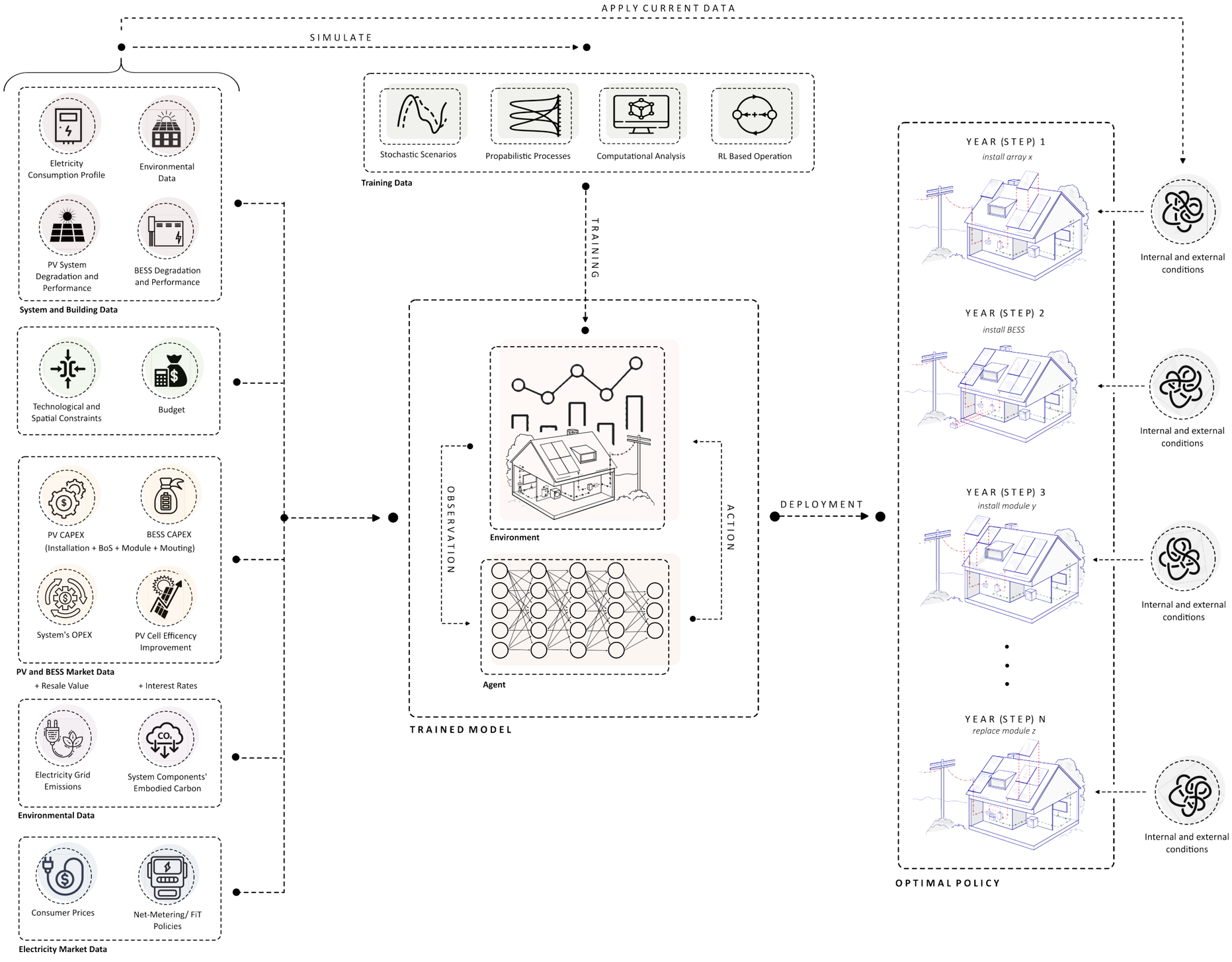


Figure 22: Workflow - Comprahensive Flowchart

3.3 General Specification

Drawing on insights from the literature review, the problem is structured as a MDP for sequential decision-making. In this framework, the agent algorithm receives observation (state) regarding the current condition of the PV or PV plus BESS setup, along with current rates of relevant variables such as electricity costs and module CAPEX. Based on these observations, the agent must execute actions aimed at maximizing the reward. The reward is calculated as the net balance between expenditures and benefits for financial optimization or as a combination of this balance with the net carbon impact associated with the installation and operation of the system.

The optimization process spans 25 years, divided into equal annual timesteps. This approach allows for a detailed analysis of long-term strategic decisions in energy system management, emphasizing both economic and environmental sustainability.

As stressed in the section 1.8, it is crucial to highlight that in steps A, B, and C, the approach is validated exclusively using synthetic data. This data is generated using the same methodology as the training data. This approach means that the trained model is evaluated solely within a controlled simulation setting and has not been tested against real-life scenarios or any external environments developed elsewhere. Such a method helps in initially assessing the model's performance but may not fully capture its efficacy in practical, real-world applications.

The roof area is split into a grid, with each grid cell being an available spot for the PV panel installation. The general constraint for the maximum number of solar panels that can be fitted on the roof (number of available gridcells) is given by the equation:

$$\sum G_i \leq \frac{(0.95 \times A_{roof})}{A_i}, \qquad \textit{eq. 33}$$

G_i is a single grid - cell,
A_i is the area of a panel.

The PV system under consideration in based on the microinverter technology since it offers the most flexibility and adaptiability in terms of dynamic changing the module configuration. The inverter is going to be chosen accordingly to panel's specification.

3.4 Utilities and Libraries

The primary Python libraries that will be utilized for the RL model are OpenAI Gymnasium, Stable Baselines 3, PyTorch.

A custom training environment is going to be built in Gymnasium. It allows for seemless development of the environment and enables seamless framework to test different algorythms (Brockman et al., 2016).

Stable Baselines 3 offers a dependable collection of reinforcement learning algorithms. Its modular design, built on the PyTorch framework, enhances its effectiveness (Raffin et al., 2021). If an issue arises, or if there's a need to test an algorithm not included in Stable Baselines 3, PyTorch will be employed for this purpose.

Additionally, Optuna will be utilized for tuning the relevant model's hyperparameters. The PV system model will be implemented using the pvlib library. Other essential libraries in mathematics and data science, such as NumPy, Pandas, and SciPy, will also be employed to support various computational and analytical needs of the project.

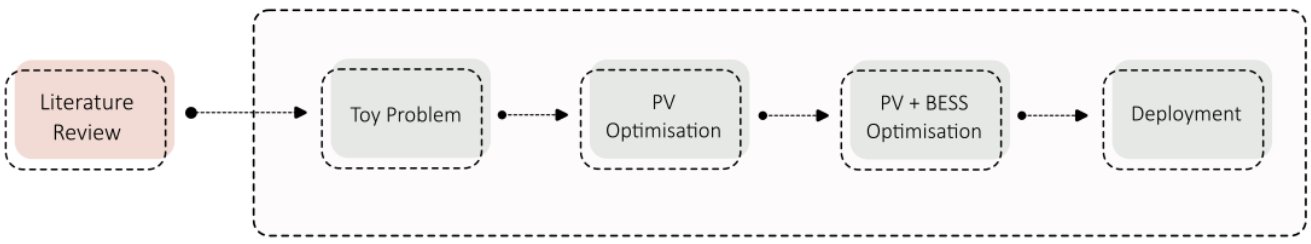


Figure 23: Workflow - Simple Flowchart

04. TOY PROBLEM

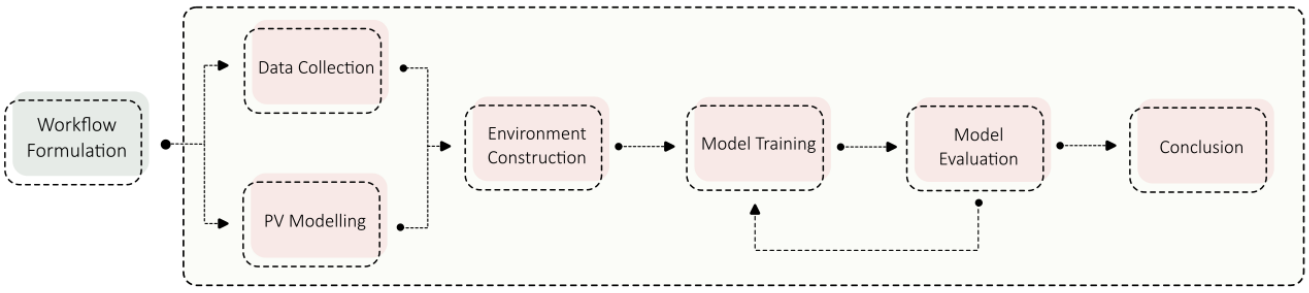


Figure 24: Toy Problem- Flowchart

04 TOY PROBLEM

This step acts as a controlled environment where the methodology can be evaluated. Here it is assumed that all relevant variables remain static throughout the optimisation period. The primary goal of this step is to test and validate the developed aproach regarding constructing the RL model environments, assess both of the selected algorythms, test and fine-tune the reward function and examine the PV system modeling process. Once the algorythm converges well, we will move on to the next stage.

4.1 PV Modelling

The annual energy output of the photovoltaic module is modelled as a single-diode using the pvlib library for python. The calculations are done using a 1-hour interval, as this is the typical frequency for weather data.

The DC output of a PV panel estimated on the I-V curve for a given timestep is calculated using the `pvlib.pvsystem.max_power_point`.

The soling conditions for the meteorological circumstances are estimated and the soling factor is included into the DC output. Then the AC output P_{AC} for a given inverter model is callculated taking into account the DC and AC cabling losses $\eta_{CabLoss}$.

The power output of a single PV panel P_{PV} for one hour is:

$$P_{PV}(h) = P_{AC}(h) \times (1 - \eta_{CabLoss}) \qquad \text{eq. 34}$$

The total energy produced by the whole array is determined by adding up the power output from each individual panel:

$$P_{total}(h) = \sum_{i=1}^{n} x_i(t) \times P_{PV}(h) \qquad \text{eq. 35}$$

where:
x_i is the binary weather a panel in a given grid cell is installed at timestep t.

The chosen module for the simulation is the SunPower SPR-MAX3-390, a residential photovoltaic panel situated in the mid to upper range category. This panel uses monocrystalline technology and possesses a nameplate capacity of 390 Wp, as detailed in Table 05.

The five parameters for the I-V curve is estimated by the `pvlib.pvsystem.calcparams_cec` given the environmental conditions and technical data from the manufacturer.

Accompanying it, the Enphase IQ7X has been selected as the integrated DC-AC microinverter, its specifications and details provided in Table 06. `Pvlib.inverter.sandia` is used to calculate the AC output.

Meteorological data utilized in this study is based on the TMY data for Rotterdam. The aggregate solar radiation on tilted surfaces is accounted by the `pvlib.irradiance` module. The cell temperature is obtained from `pvlib.temperature.pvsyst_cell`.
As anticipated, losses due to soiling are marginal. The calculated, summarised daily DC and AC outputs for a single module are depicted in Figure 25.

This study assumes the DC and AC cabling losses $\eta_{CabLoss}$ as the mean value calculated by Hashemi et al. (2021) at 1%. The surface albedo has been assumed 0.25, for an urban environment. The total electricity produced over one year ammounts to 4040 kWh, not bad.

4.2 Energy Balance

Thus, for any hour h, if $P_{total}(h) < E_{load}(h)$ the hourly energy supply from the utility grid is formulated as $E_{sup}(h) = E_{load}(h) - P_{total}(h)$; the total annual energy supplied from the grid $E_{sup}(h)$ is:

$$E_{sup}(t) = \sum_{h=1}^{8760} E_{sup}(h) \qquad \text{eq. 36}$$

On the other hand, for any hour h, if $P_{total}(h) > E_{load}(h)$ the hourly excess energy exported to the grid is $E_{exp}(h) = P_{total}(h) - E_{load}(h)$; the total annual energy exported to the grid $E_{exp}(h)$ is:

$$E_{exp}(t) = \sum_{h=1}^{8760} E_{exp}(h) \qquad \text{eq. 37}$$

The electricity consumption load has been modelled synthetically on the basis on a single - family terraced house (Appendix B) constructed prior to 1945 (Wahi et al., 2023). This particular house serves as a representative example of its typology in terms of its electrical consumption patterns. It features a roof alligned along the north-south axis. It is presumed that the roof and electrical instalation is suitable for PV installation.

4.3 RL Environment

4.3.1 General Specification

The optimal planning for PV panels can be formulated as an MPD described by the five-tuple:

$$M = (S, A, P, R, \gamma)$$

In each timestep t, the environment determines the impact of the agent's actions $A(t)$ on it and how the current state of the environment $S(t)$ influences the agent. The agent will continually refine the policy until

it achieves convergence. It gathers state information from the environment, such as current variable pricing, electricity grid (EG), and rooftop PV configuration.

4.3.2 Action Space

The action space $A(t)$ represents the set of all possible actions for the PV system state. The set of discrete actions is chosen to most effectively embody the possible step that the owner might take concerning each of the PV panels at timestep t. Those are defined as:

Preserve – the agent makes no change to the grid cell G_i

Install – the agents adds a panel to an empty gird cell G_i

Discard – the agent removes a panel from the grid cell G_i

Replace – the agent replaces a panel from the grid cell G_i

4.3.3 Observation (State) Space

The observation space defines the state the agent can observe. It is composed of local variables – specific for each grid cell G_i (the binary installation state x).

$$OBS(t) = [x_i, x_{i+1}, x_{i+2}, \dots, x_{i+n}] \quad \textit{eq. 35}$$

4.3.4 Reward Function

The fundamental premise of DRL is to maximize reward, making the reward function a crucial element. An effectively crafted reward can improve learning efficiency and hasten the convergence process (Sutton & Barto, 2020). The economic gains, losses and the carbon emissions offset from the grid incurred in each step are determined according to the observed space and the chosen action. It is clear that the goal here is to maximise the overall profit from the PV system utilisation and to minimise the carbon emission thrughout the entire training episode.

The annual financial balance is defined as:

$$BalanceFIN(t) = \{[-C(t) - I(t) \times E_{sup}(t) + J(t) \times E_{exp}(t)] - Pen(t)\} \times (1-\vartheta)^t \quad \textit{eq. 36}$$

where:
$C(t)$ is the annual total cost of the PV system ($CAPEX(t)$ + $inverter(t)$ + $OPEX(t)$),
$I(t)$ represents is electricity tariff at year t,
$J(t)$ represent the compensation tariff at year t,
ϑ is the discount factor,
$Pen(t)$ is the penalty for exceeding the maximum annual budget.

The last part of this equation ϑ is a discount rate adjusting the time value of money. This means that the annual balance at the start of the 25 year long optimisation period is "worth" more than the balance at the last timestep. This is done to account for inflation and for the increasing uncertainty of the future balances. ϑ is equal to 4%.

On the other hand $Pen(t)$ is defined as:

$$if\ C(t) \le B(t) \quad Pen(t) = 0 \quad \textit{eq. 37}$$

$$if\ C(t) > B(t) \quad Pen(t) = [B(t) - C(t)] \times (1 + r)$$

where $B(t)$ is the annual budget, r the annual loan interest rate, which is fixed at 6%. This rate aligns with the standard interest rate for green loans in the Netherlands, which are utilized for household renovations and energy efficiency enhancements. The penalty accounts for a realistic scenario where the household has a limited ammount of money to spent each year and going over that limit results in taking a loan. A more detailed explantion is provided in Appendix H.

Luckily, the environmental balance is much simpler:

$$BalanceEN(t) = [E_{tot}(t) \times EG(t)] \quad \textit{eq. 38}$$

where:
$Etot(t)$ is the total energy produced by the PV panels at t,
$EG(t)$ represents the grid emission factor at t.

Hence the cumulative reward for each training episode can be expressed as:

$$R = \sum_{t=1}^{25} BalanceFIN(t) + \left[w \times \sum_{t=1}^{25} BalanceEN(t) \right] \quad \textit{eq. 39}$$

w is the weight parameter to calibrate the value of both total balances, it is going to be set and adjusted later.

4.3.5 Environment Building and Action Space

As described in section 3.2.2. the training environment has been built in Open AI Gymnasium, as an object of the `gym.Env` class. The action space is constructed as

a multidiscrete space with integers ranging from 0 to 2.

The action space is configured as following:

0 – preserve - the agent makes no change to the grid cell G_i,

1 - install – if the observation was 0 in the previous timestep or replace - if the observation was 1,

2 - discard – if the observation in the previous step = 1 the agent removes a panel from the grid cell G_i, else do nothing.

$$A = \{[a_i, a_{i+1}, a_{i+2}, \dots, a_{i+n}]\} \quad \textit{eq. 40}$$

Where a_i represents the action assigned to each panel, which is an integer value of either 0, 1, or 2. The reward function is set up and both total balances are normalised. The weight parameter w, is assigned a value of 3. This adjustment ensures that the average financial balance accumulated over an episode is given three times more importance than the total environmental balance in the calculation of the final reward. This reflects the supposition that financial considerations are likely to be of greater importance to a residential user than the CO_2 footprint.

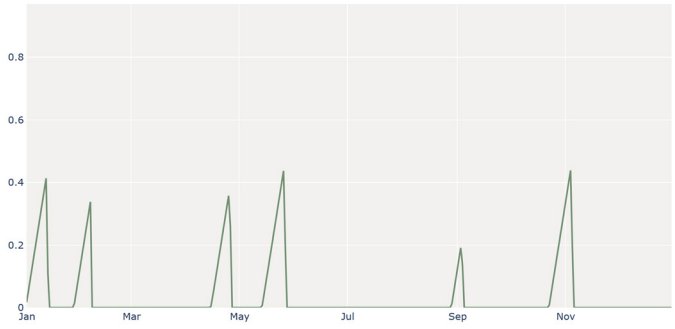


Figure 25: Soiling Build-up as % of radiation losses

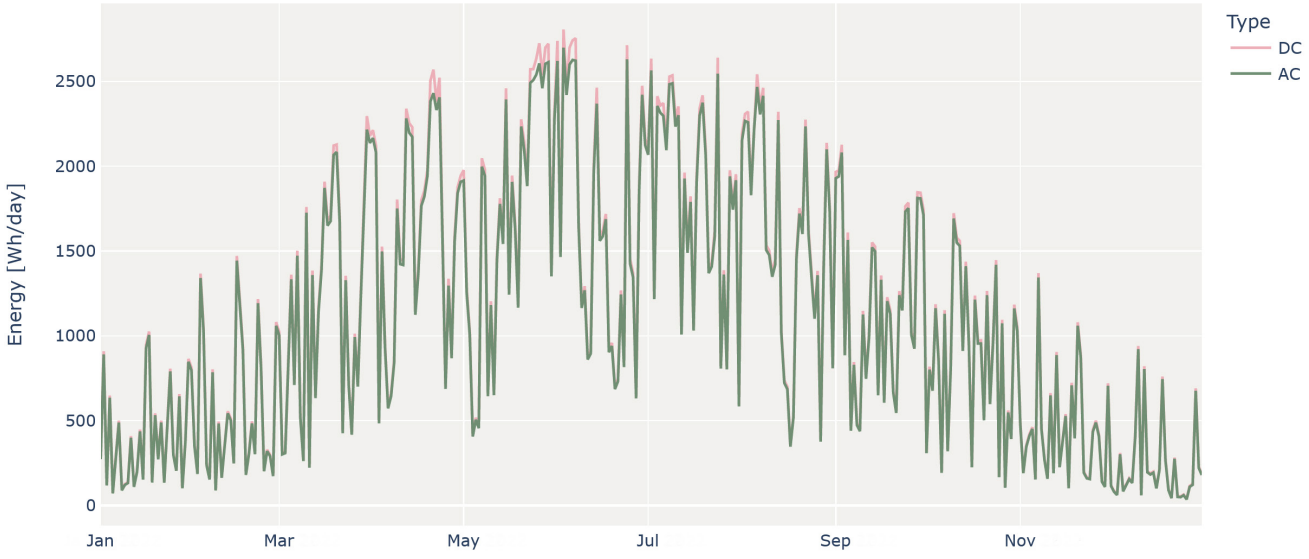


Figure 26: Annual output of a single module

P_{max}	A	N_s	I_{sc}	V_{oc}	I_{mp}	V_{mp}	α_{sc}	β_{oc}	a_{ref}	I_L	I_o	R_s	R_{sh}	γ_r
390 Wp	1.77 m^2	104	6.55 A	75.3 V	6.05 A	64.5 V	0.0025	-0.199	2.631	6.558	$1.37e^{-11}$	0.309	268.7	-0.351

Table 06: PV module specification

V_{ac}	P_{so}	P_{dco}	P_{aco}	V_{dco}	C0	C1	C2	C3	P_{nt}	$V_{dc(max)}$	$I_{dc(max)}$
240 V	1.3	315		60	-0.0...28	-0.0...25	-0.0034	-0.0378	0.021	64	5.3828

Table 07: Inverter Specification

4.4 Variables

As mentioned before all relevant variables are assumed to be static.

The climatic conditions for energy production are modeled under the typical meteorological year data for Rotterdam. No degradation of the system is taken into account.

The energy tariff is assumed flat throughout the day. and along with CAPEX and EG (grid factor) is constant, equal to their initial values in 2024 at $t=1$. PV cell efficency does not improve over time. The net-metering rate is assumed to be 80% of the electricity tariff. OPEX is assumed to be 2% of annual CAPEX.

Cost data was sourced from the official websites of manufacturers and retailers. The data concerning the carbon emissions for a single panel was estimated based on the average carbon dioxide equivalent per square meter for a silicon photovoltaic panel on the European grid, as reported by Wikoff, Reese, and Reese (2022).

Consistent with the majority of LCA research on photovoltaic technology, this study excludes the consideration of end-of-life recycling and the carbon footprint of the microinverter due to the absence of commercially established processes and available data.

All variables needed for the reward calculation, with sources are given in the table 08.

Variable	Value	Source
Max. no of panels (gridcells G_i)	8	-
Electricity Price	0.475 €/kWh, Fixed,	Eurostat, 2024
Net-metering	0.475 €/kWh, equal to the electricity tariff	SolarPower Europe, 2023
Panel Cost	346 €	Europe-Solarstore, 2024
Microinverter Cost	170 €	Enphase, 2024
BoS and instalation costs	85% of the panel costs, if more than one panel is installed it decreases by a factor of by 0.1 for each installed module.	IRENA, 2022
Total CAPEX	673.2 € if one panel installed	-
OPEX	2% of the combined module and inverter costs	-
Electricity Load	111 kWh mean hourly use	Wahi, 2023
Panel Carbon Footprint	424.8 kgCO_2	Wikoff, Reese and Reese, 2022
Surface Albedo	0.20 as a mean for an urban area	Zhang et al., 2022

Table 08: Variable Values

4.5 Evaluation Metrics

The final outcome of both algorythms shall be evaluated in order to assess which algorithm, in what hyperparameter configuration performs better. Three evaluation approaches are defined. The first one is the cumulative reward gained by the model on the validation dataset.

$$C_{tot} = \sum_{t=1}^{T} C(t) \qquad \textit{eq. 42}$$

The second one, is defined by the internal rate of return (IRR) that indicates the cumulative annual return on investment. This will also help with the evaluation of the reward function. It is the discount rate that makes the net present value (NPV) of all cash flows from a particular project equal to zero. In other words, the IRR is the annualized effective compounded return rate that can be earned on the invested capital (Khezri et al.):

$$-C_{tot} + \sum_{t=1}^{T} S(t) \times (IRR)^t = 0 \qquad \textit{eq. 41}$$

where:
$S(t)$ is the net cash flow at t,
C_{tot} is the total value of the investment.

On the other hand to evaluate the environmental benefit of the PV system life cycle assessment can be conducted. Here, the cradle-to-cradle CO_2 emissions are determined for the PV module. The value of emission is expressed as kg CO_2 eq/module. The CO_2 emissions are divided by the produced electrical energy to compute the system's Carbon Intensity (CI):

$$CI = \frac{CO_{2,tot}}{E_{tot}} \qquad \textit{eq. 43}$$

where:
$CO_{2,\,tot}$ is the total equivalent CO2 emission of the system.

4.6 Training and Results

Two versions of the toy environment were developed. Version 1. without any budget constraints or penalties and version 2. with a complexity of a budget constraint as described in section 3.3.6.

It has been discovered that DQN is incapable of managing multi-discrete action spaces, as utilised in this study. Consequently, in addition to PPO, another DRL algorithm, Advantage Actor-Critic (A2C), is being considered for comparison and evaluation.

The hyperparameters were iteratively tuned through a trial and error process. The final configurations of the hyperparameters are detailed in the following section.

4.6.1 PPO

The hyperparameters set for the training of the PPO agent are outlined below. Learning rate is set to 0.001 at the start and decreases lineary to 0.00001. PPO's policy is a simple neural network. The discount factor, gamma, was set to 1 due to the nature of the task, which has a definitive end point within a finite number of steps. The lambda parameter in Generalized Advantage Estimation (GAE) serves to balance the bias-variance trade-off in value function estimation by controlling the weighting of multi-step returns. It was set at 1, since the reward is given only at the end of an episode. The numbers of epochs, batches, and interations were set at 10, 64, and 300 000 respectively. Both the actor and critic networks feature two layers, each with 64 neurons, and use the Tanh activation function.

At the end the cumulative reward is divided by a factor of 350, as it has been observed that the agent achieves faster convergence when the reward magnitude is nearer to 0.

4.6.1.1 No constraints scenario - Results

The outcomes of the training align with expectations, indicating that implementing action 1 at the initial timestep for all grid cells is the optimal strategy, as there is no degradation or compromise involved in early panel installation. Subsequently, the model consistently selects action 0 for the remaining timesteps until the episode concludes. The total financial balance is -17 959 € compared to -49 535 € when no panels are installed. The achieved IRR is 71% - very high. The system manages to save over 32 100 kg CO_2 eq. over its lifespan. The carbon intensity is 0.593 kg CO_2/kWh.

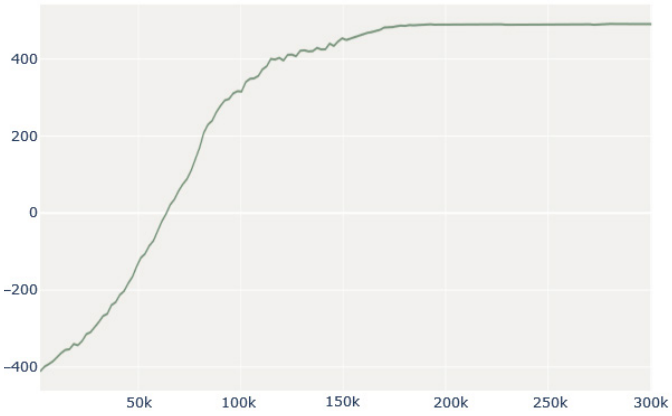


Figure 27: PPO no budget training

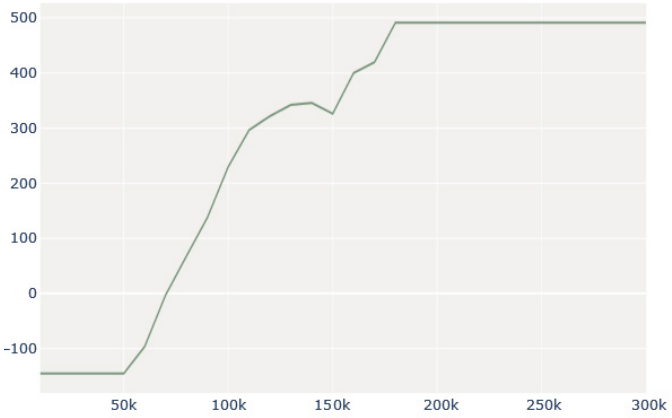


Figure 28: PPO no budget evaluation

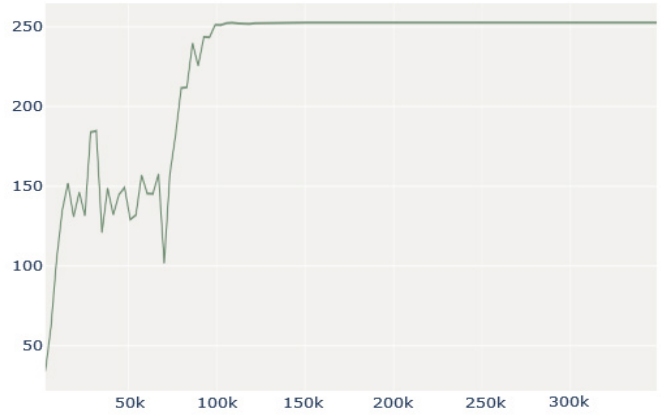


Figure 29: A2C no budget training

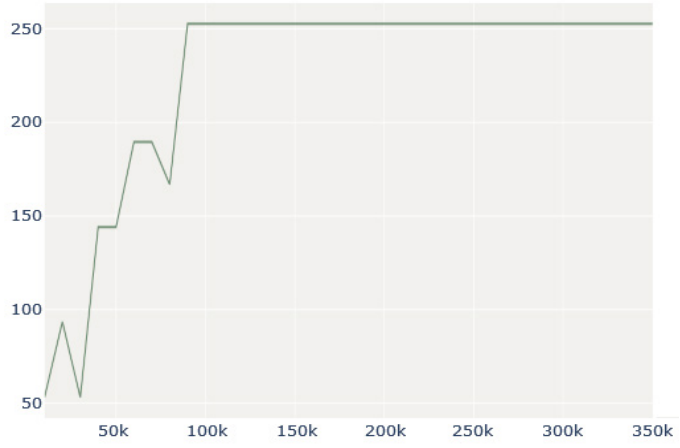


Figure 30: A2C no budget evaluation

4.6.1.2 Budget constraint scenario - Results

Incorporating a budget constraint into the environment introduces additional complexity, leading to a degradation in outcomes. Convergence occurs slower. Depending on hyperparameters, the agent's ability to consistently identify the optimal policy is slightly compromised, resulting in sporadic, suboptimal decisions, such as the addition of extra panels to the already filled grid.

Compared to the constraint-less scenario the optimal convergence has been achieved using a fixed learning rate of 0.0003 and 32 neurons in each NN layer.

After 1000 evaluation episodes, the average financial balance stands at -€18,831, an improvement over the -€49,535 observed without panel installation. The Internal Rate of Return (IRR) achieved is 43%. The system contributes to a CO2 equivalent saving of more than 32,071 kg throughout its operational life, with a c of 0.598 kg CO2/kWh.

4.6.2 A2C

Training the A2C agent proved to be significantly more challenging than training with the PPO algorithm. Initial attempts often led to issues such as exploding gradients, indicating instability in the training process. Overall, in comparison to PPO, the A2C training exhibited lower stability and slower convergence rates.

Learning rate is set to 0.0007. Gamma and lambda factors are set to 1. The value function coefficient and maximum gradient norm are both 0.5. Both the actor and critic networks feature three neural network layers, with 64 neurons each and ReLU activation functions.

4.6.2.1 No constraints scenario - Results

The model fails to select the action 1 for all gridcells at timestep 1. It covers 4 out of 8 cells, selecting the action 0 for the rest of steps within the episode.

The total financial balance is -32694 € compared to -49 535 € when no panels are installed. Similarly to PPO, the achieved IRR is 71%. The system manages to save over 12 037 kg CO_2 eq. over its lifespan.

4.6.2.2 Budget constraint scenario - Results

Here, the mean total financial balance for 1000 evaluation episodes accounts to -€22,121. The system manages to save over 20 000 kg CO_2 eq. over its lifespan. IRR is 44% and gwp 0.593 kg CO_2/kWh.

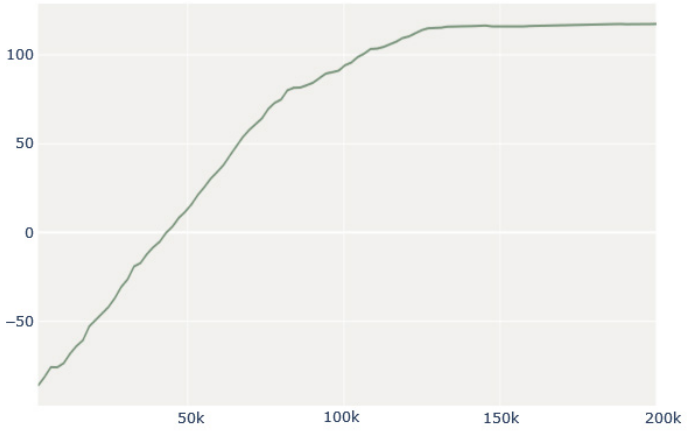


Figure 31: PPO budget training

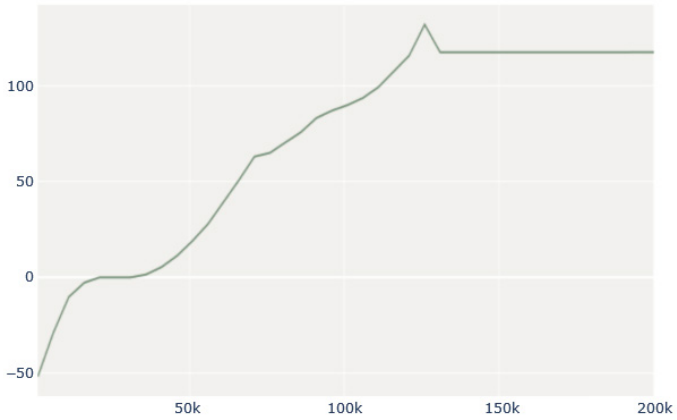


Figure 32: PPO budget evaluation

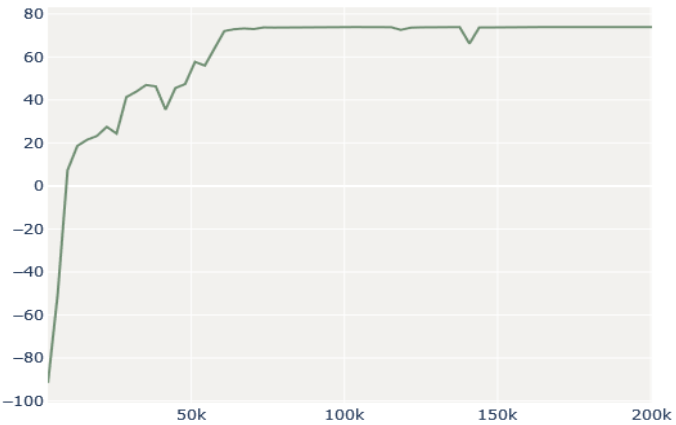


Figure 33: A2C budget training

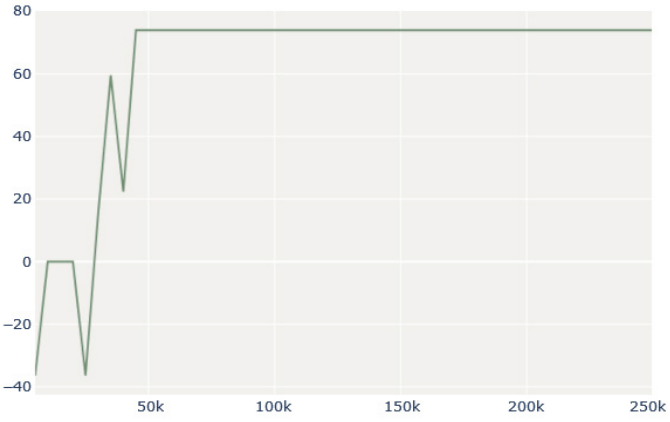


Figure 34: A2C budget evaluation

4.7 Conclusions

In this phase, a training environment was established, and various models were systematically trained within it. The initial evaluation of the TOY problem indicates a successful outcome, as the models demonstrated effective convergence, yielding anticipated results. This achievement authorizes the progression to the subsequent phase of the project. Two algorithms, Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C), were tested and assessed across two distinct scenarios and a spectrum of hyperparameters.

It was observed that the agents generally encountered increased difficulty when operating under budget constraints. Comparative analysis reveals that the PPO algorithm outperforms A2C in terms of convergence speed and reliability of results.

It is important to note that the use of subprocesses, specifically leveraging multiple CPU cores to train the environment in parallel, was not incorporated in this section. Implementing parallel processing could potentially enhance the convergence speed of the algorithms by distributing computational tasks across several cores, thereby reducing the time required for the model to reach optimal solutions.

05. PV System Size Optimisation

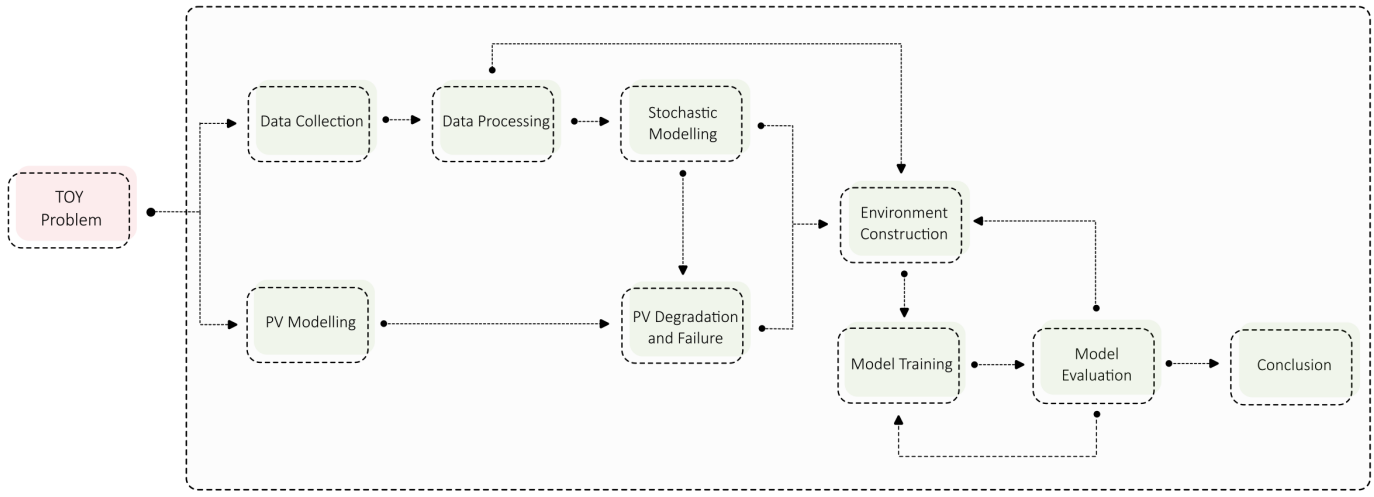


Figure 35: PV system Ooptimisation - flowchart

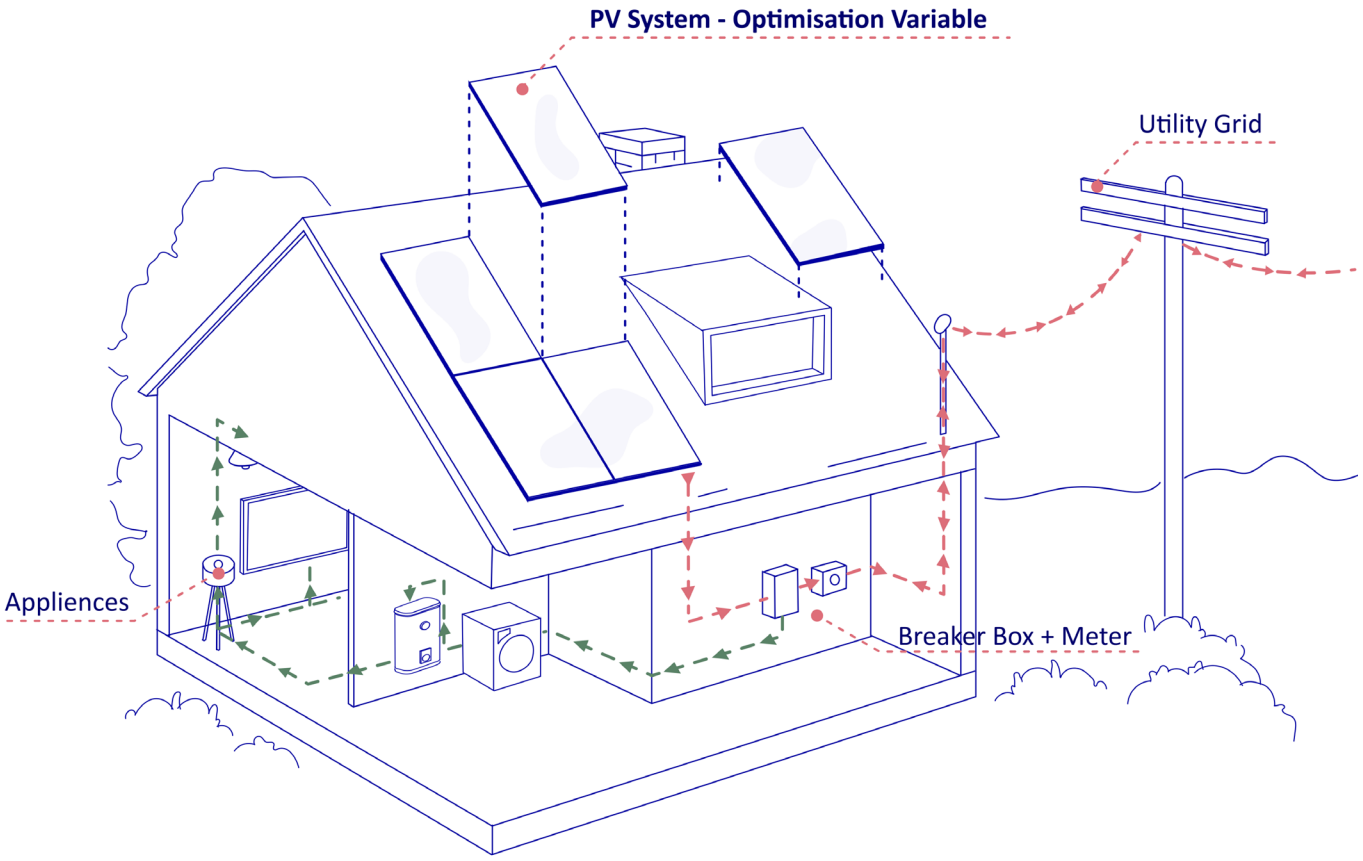


Figure 36: PV System Size Optimisation

05 PV OPTIMISATION

After completing and assessing the toy problem, the experimental process progresses to the Full-Scale Development stage. Here, the complexity of the variables increases gradually. Each variable is introduced individually, followed by an evaluation of the algorithm's performance and output. The methodology from the previous chapter for building the environment, rewards, states and actions is considered a baseline approach here.

To validate the developed approach, it is tested on a wide range of possible scenarios. Three base panels have been chosen for analysis. These were selected to represent various economic tiers as well as the range of technologies currently accessible in the consumer market. Furthermore, two distinct annual budget ranges have been identified, subject to adjustment based on the annual interest rate in subsequent steps. Additionally, three house loads are selected, with varying roof inclinations and sizes.

Each combination of these conditions will be incorporated into separate RL environments. An algorithm will be trained on each of theese configurations, to validate the sensibility of the proposed methodology.

First the optimisation is focused on the financial aspect only. Then the environmental aspect is going to be introduced.

5.1 PV Modelling

The methodology for modeling the panel's output aligns with a toy problem approach, additioanlly incorporating both the annual degradation rate and the survival rate to simulate the panel's failure.

In addressing a financial and environmental optimization within the context of PV system design, three distinct solar panel and inverter configurations were selected. This was informed by a desire to encompass a comprehensive range of price levels, technological attributes, and environmental impacts. The inclusion of configurations that represent high-end, budget, and used categories allows for the evaluation of cost-effectiveness.

Technologically, the chosen configurations embody a spectrum of solar panel technologies—namely, monocrystalline N-type, PERC (Passivated Emitter and Rear Cell), and polycrystalline. This variety facilitates a comparative analysis of performance efficiencies, maintenance requirements, and longevity, offering insights into how technological differences impact both upfront costs and long-term returns.

In terms of environmental considerations the inclusion of a used panel configuration which carries a nominal carbon footprint provides a perspective on the reuse and lifecycle implications of photovoltaic technologies.

Base Model and Inverter	**Maxeon SPR-MAX6-435 + Enphase IQ8HC**	**JA Solar JAM54S30-415/ MR + Enphase IQ8+**	**TrinaSolar Module TSM-265 PC05A + Enphase M250**
Tier	Upper - Range	Budget	Second-Hand
Technology	Monocristaline, N-type, IBC	Monocristaline, PERC	Polycristaline
Module Efficency	23.3%	21.3%	16.2%
φ Weibull parameter	50	30	15
CAPEX (inverter included)	1,583 €/Wp	0,69 €/Wp	0,36 €/Wp
Total DC Output at t =1 for House 1	495618 Wh	463464 Wh	292251 Wh
Total AC Output at t =1 for House 1	468343 Wh	438051 Wh	284870 Wh
Mean Degredation Rate (μ gaussian parameter)	0.67 %	0.82 %	1.19 %

Table 09: Overview of the Selected Models

Specifications are presented in table 9. More detailed parameters are in Appendix G.

5.1.1 Degradation Rate

At the time of writing pvlib does not account for the long-term efficiency losses. Following Lai et al., 2023 gaussian probability distribution with constant variance was applied to the annual degradation rate of the PV modules.

The degradation of photovoltaic modules is affected by irradiance and, particularly, by the temperature of the PV cells. Research by Nehme et al. (2020) suggests that a decrease of 1 degree Celsius in cell temperature can result in up to a 2.5% reduction in the rate of efficiency loss over time periods where mean temperatures exceed 40 degrees Celsius. Conversely, Rahman et al. (2023) contend that factors such as humidity and soiling play a more significant role in the efficiency degradation of PV systems than temperature variations.

In this study, despite uniform meteorological conditions across all cases, variations in cell temperature may still arise due to different roof inclinations and orientations. Investigations into cell temperature under these specific conditions have revealed that mean PV cell temperature variations are minimal, ranging from 15 to 17.4 degrees Celsius. Given these findings, it is reasonable to assume that the degradation conditions for all cases analysed are equivalent.

Degradation may exhibit a slight temporal dependency, potentially increasing marginally over the lifespan of the system. However, given the limited availability of long-term data—particularly over durations as extended as 25 years —it is assumed that the mean degradation rate remains constant.

In this study the degradation values for mono- and poli-crystalline modules located in temperature zones 2 and 3 were extracted from a publicly accessible dataset provided by the National Renewable Energy Laboratory (see Appendix E) and fitted into gaussian distributions for each base model according to its technology.

To make the data more closely align with a normal distribution, extreme values were excluded from the analysis and square root transforms were applied. To check the goodness of fit Kolmogorov-Smirnov (KS) test was perfomed with the resulting p-values ranging from 0.756. to 0.011.

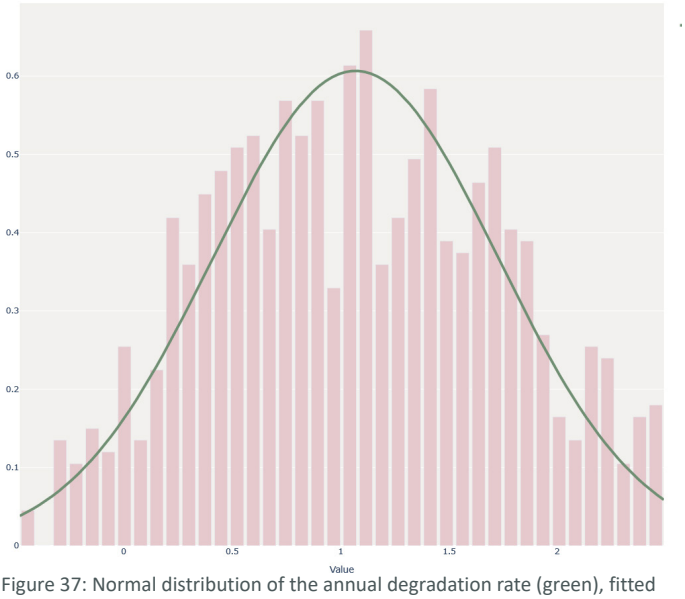


Figure 37: Normal distribution of the annual degradation rate (green), fitted to the data (pink) for the budget panel.

In the environment, the degradation rate for a single panel is randomly sampled from the normal distribution and then substracted from the panel's efficency.

$$\eta_i(t) = \eta_i(t) - \epsilon_t \qquad \textit{eq. 44}$$

Figure 33 displays an exemplary distribution of the data, with a fitted normal distribution curve overlaid to illustrate the underlying statistical properties.

5.1.2 Survival Rate

The **survival rate** of PV panels is represented using the Weibull distribution, chosen for its simplicity and established use in reliability engineering. The survival function for individualised each panel i which is the complement of the cumulative density function can be expressed as:

$$R_{PV,i}(t_i) = e^{-(\frac{t_i}{\varphi})^{\beta}} \qquad \textit{eq. 45}$$

where:
$R_{PV,i}(t)$ represents survival rate of an ith panel in the year t,
t_i is the time from the installation of a given panel to the current timestep t when the survival rate being evaluated ($t_i = t - t_{install,i}$),
φ is the scale parameter,
β is the shape parameter of the Weibull law.

This study assumes $\beta = 3$ (Fan & Xia, 2017) which indicates a wear out failure. The φ parameter are set based on available literature on each model (Appendix F).

To indicate whether the specific panel is operational at the year t, a parameter θ_i is randomly sampled for each

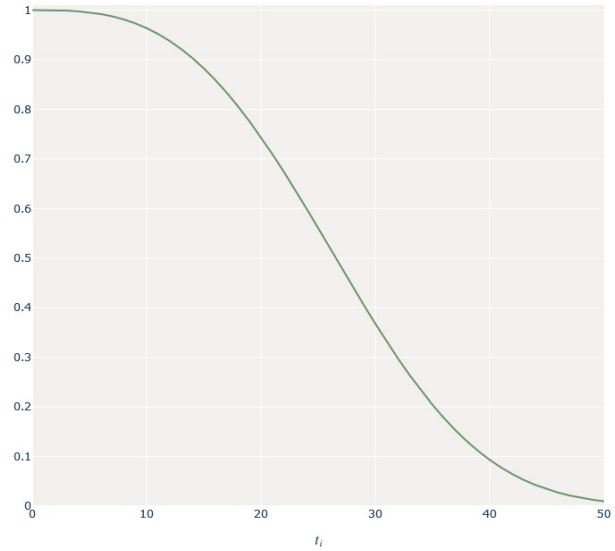


Figure 38: Weibull distibution for the Budget panel.

panel from the Weibull probability density function. This can be represented by a binary variable $W_{PV,i}(t)$ for each panel i at each timestep t:

$$W_{PV,i}(t) = \begin{cases} 1 & if\ \theta_i \leq R_{PV,i}(t_i) \\ 0 & if\ \theta_i > R_{PV,i}(t_i) \end{cases} \qquad eq.\ 46$$

Hence the power output of a single PV panel P_{PV} is for one hour given by:

$$P_{PV}(h) = P_{AC}(h) \times W_{PV,i}(t) \qquad eq.\ 47$$

The total energy produced by the whole array is determined by adding up the power output from each individual panel:

$$P_{total}(h) = \sum_{i=1}^{n} x_i(t) \times P_{PV}(h) \qquad eq.\ 48$$

5.2 Building Loads

Three different house loads were selected for the simulations

House 1 is a typical terraced house in Delft with three facades, built in 2004/2005. Electricity usage includes lighting, standard household appliances, electric cooking, MVHR, and an electric hot water boiler. The boiler is pre-heated by block heating (varying in temperature between 25°C and 50°C depending on the outdoor temperature).

Hourly data spanning three consecutive years (2018-2020) was collected and subsequently averaged.

The total annual electricity consumption ammounts to 3,045.2 kWh a year which is slightly less than the Dutch average for a single-family house for 2022 (3,090 kWh).

Due to the lack of more loads from the Netherlands immediately available, house loads 2 and 3 have been sourced from an open dataset described in Schlemminger et al., 2022 from north-west Germany. Both were built in late 90s and early 2000s and are equipped with water-water-heat pumps connected to a cold local heating network and solar thermal systems for domestic hot water.

Parameters for each house are detailed in Table 10. Note that the data pertaining to roof size, orientation, and inclination are hypothetical.

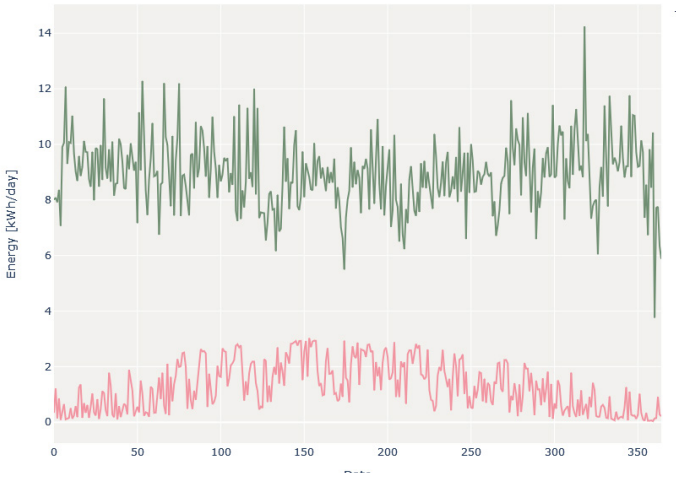


Figure 39: Plot of the daily summarised electricity consumptionof house 1 (green) and Maxeon panel (pink)

5.3 Electricity Tariff

For simulating the scenarios for the consumer electricity tariff GBM will be used. The benefits of using brownian motion driven stochastic processes for producing monte carlo simulations of electricity tariff was described in section 2.4.2. GBM was selected due to its straightforward implementation and well-established presence in the literature. Having obtained the past data, log returns are computed using the formula:

$$logX_t = ln\left(\frac{X_t}{X_{t-1}}\right) \qquad eq.\ 50$$

The volatility parameter σ can be estimated as the

Variable	Sigma	Mu	Theta
Annual Electricity Tariff	0.068	0.031	-
Grid Factor	0.016	0.036	0.067
CAPEX	0.031	-0.024	-

	Lambda	-	Theta
Cristaline Cell Eff. Advancement	3.725		0.706

Table 11: Parameters of the stochastic processes

House	1	2	3
No of Inhabitants	2	2	3
Type	Terraced	Detached	Detached
Built	2004/2005	early 2000s	early 2000s
Square Footage	unknown	135	203
Roof Type	Single-pitched	Single-pitched	Double-pitched
Surface Albedo	0.20 (Zhang et al., 2022)	0.20 (Zhang et al., 2022)	0.20 (Zhang et al., 2022)
Roof azimuth and tilt	180 ° and 40 °	180 ° and 30 °	90 °/ 270 and 30 °
Max no of modules	12	24	36 (18 each side)
Annual Load	3045 kWh	3836 kWh	4678 kWh

Table 10: Overview of the analysed houses

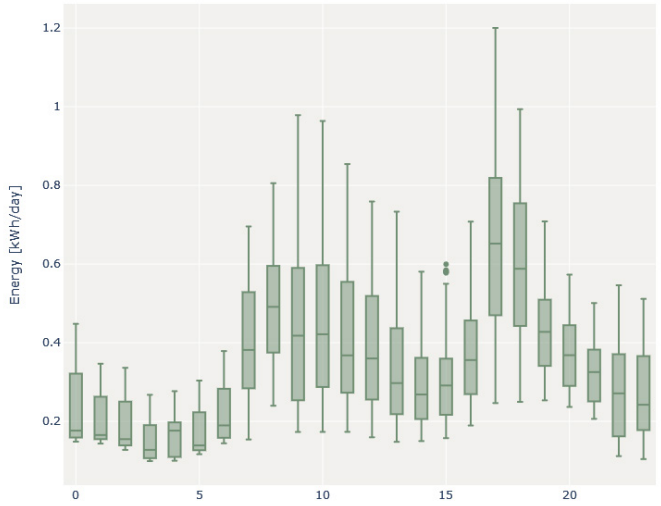


Figure 40: House 1 Daily Load Box Plot

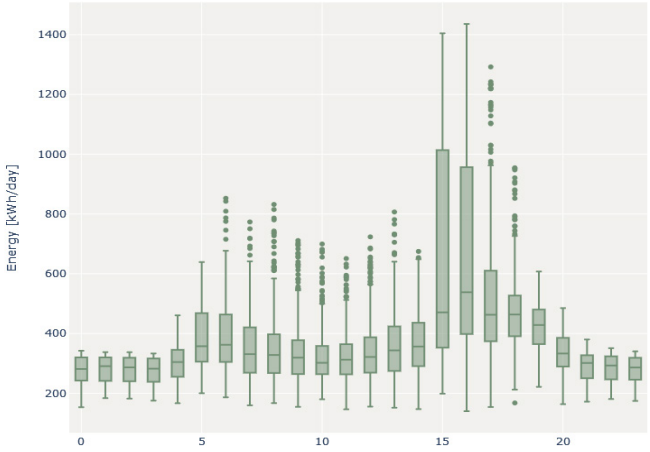


Figure 41: House 2 Daily Load Box Plot

standard deviation of log returns:

$$\sigma = \sqrt{var(logX_t)} \qquad \textit{eq. 51}$$

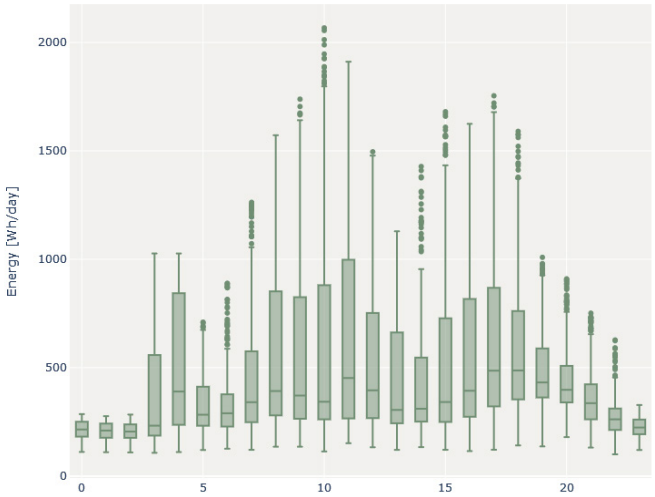


Figure 42: House 3 Daily Load Box Plot

where:
n is the total number of log return observations.

Then, the drift parameter μ normalized over a time interval that corresponds to one year, can be estimated as the mean of the log returns adjusted for the drift:

$$\mu = mean(logX_t) \times \frac{\sigma^2}{2} \qquad \textit{eq. 52}$$

Historical data on residential electricity tariff was obtained with annual timestep (Appendix C). First both parameters were estimated. 10000 paths were simulated in total (fig. 43). The tariff at t = 0 is assumed to be the the average condumer price in the Netherlands in January 2024 (Rijksdienst voor Ondernemend Nederland, 2024). The bold red line indicates annual means.

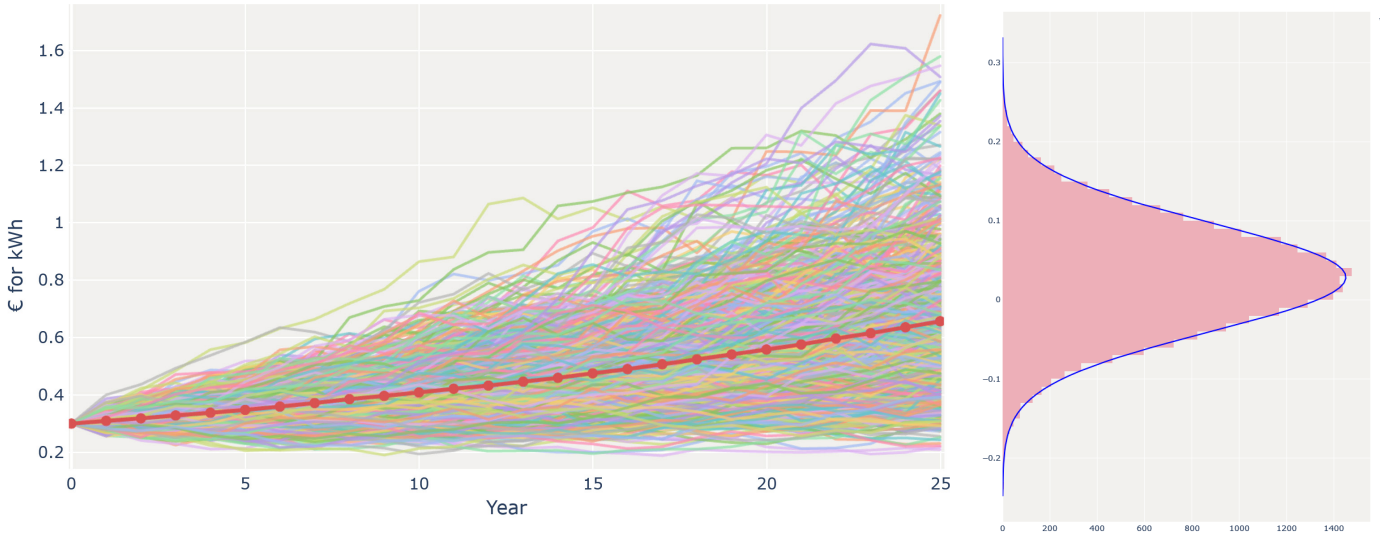


Figure 43: Simulated Scenarios for the Annual Electricity Tariff and the distribution of the random steps

5.4 Grid Factor

As expressed in part 2, no describtion of stochastic modelling of EG can be found in literature. Yet, it can be determined that GBM, which assumes a constant relative and unbounded growth rate might not be the suitable solution. Grid emission factors are often influenced by regulatory policies, technological advancements, and market mechanisms, all of which work towards maintaining a balance or target level (Tsiropoulos et al., 2020). In the Netherlands for instance (Scheepers et al., 2022), the long term goal is set for the 95% emission reduction by 2050 compared to 1995.

Mean-reverting processes, such as the Ornstein-Uhlenbeck (OU) process, are advantageous for modeling quantities that tend to return to a long-term mean, thus avoiding extreme long-term forecasts. This process is considered suitable for the grid factor, based on the assumption that once the target grid emissions level is achieved, the emissions will fluctuate around a specific mean with only minor deviations. This implies that emissions will neither increase significantly nor fall below zero.

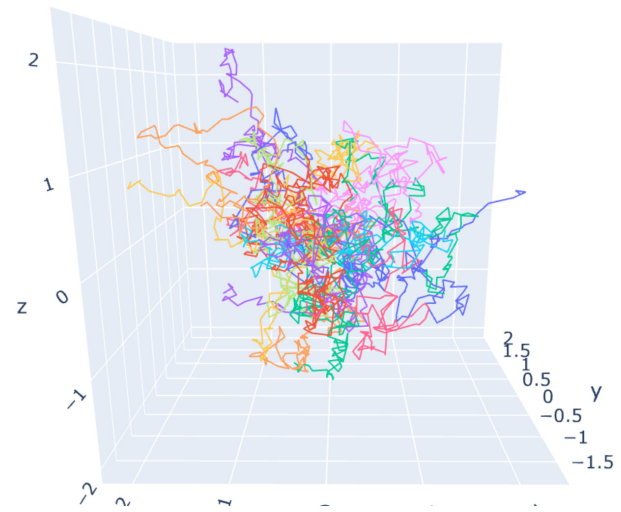


Figure 44: Simulation of 3-D Brownian Motions 25 Steps Each

Similarly to GBM OU can be expressed explicitly as (Asango, 2018):

$$X_t = X_0 e^{-\lambda t} + \mu\left(1 - e^{-\lambda t}\right) + \sqrt{\frac{\sigma^2\left(1-e^{-2\lambda t}\right)}{2\lambda}} N(0,1)$$

X_t denotes the EG at time t, *eq. 53*
μ is a long-term mean coefficient,
λ is a mean reversion rate,
σ is the volatility of the process,
X_0 is the EG at the start of the process.

When X_t exceeds the asymptotic mean μ, the drift becomes negative, drawing the process towards the mean. Conversely, when X_t is below μ, the drift acts in the opposite manner.

To estimate the parameters of the OU process historical data was used to calculate the process' volatility, while its long-term mean and reversion rate were estimated from projected CO2 equivalent emissions in the Dutch grid system, extending up to the year 2050. The parameters are calculated with maximum likelihood estimation (MLE) using a dedicated *ouparams* library.

5.5 Net-Metering

Under the circumsances described in 2.4.4 the compensation tariff $J(t)$ in 2024 (*when t = 0*) is equal to:

$$J(t=0) = I(t=0) \qquad \textit{eq. 54}$$

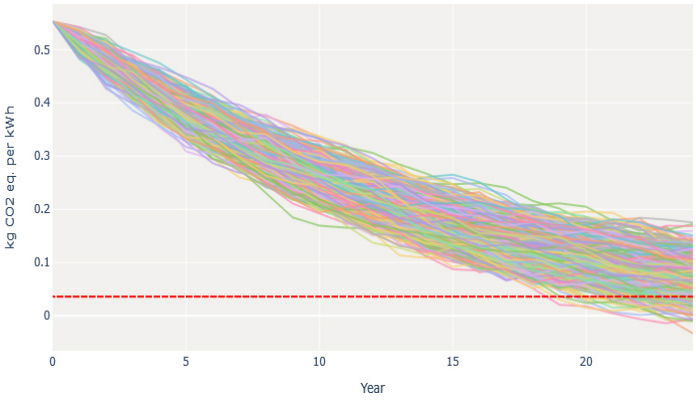


Figure 45: Simulated Scenarios for the Grid Factor (long term mean - mu marked in red)

and in year 2029 is equal to:

$$J(t=5) = I(t=5) \times 0.37 + I(t=5) \times 0.8 \times 0.63$$

etc., *eq. 55*

where:

$I(t)$ is the electricity tariff at year t.

The feed-in-tariff replacing the net-metering scheme is given as 1/3 of the electricity tariff (Norouzi et al., 2023).

5.6 Cristaline Cell Development

The primary determinant of technological advancement in photovoltaic technology is the efficiency of photovoltaic cells. It is a metric that quantifies the proportion of incident solar radiation that a PV cell can convert into electrical energy. It is expressed as a percentage, reflecting the effectiveness of the cell in harnessing solar energy under standard test conditions.

$$Efficency_{cell(STC)} = \frac{P_{max}}{1m^2 \times P_{in}} \times 100\% = \frac{I_m \times V_m}{1m^2 \times P_{in}} \times 100\%$$

where: *eq. 56*

$P(m)$ is equal to 1000 W/m^2

From equation 56 it is evident that there is a linear relationship between cell efficency and the maximum power output that the panel can produce under standart test conditions (P_{max}). Therefore, this study assumes linear proportionality between the efficiency progression of crystalline cells over time and the power output of new photovoltaic panels available on the market.

The historical data on conversion efficiencies for best research cells was obtained from National Renewable Energy Laboratory (2024) (Appendix I).

The data indicates a pattern of a gradual improvement over time, marked by intervals of consistency followed by jumps in efficiency. The monte carlo scenario generation is conducted using a discretely sampled Gamma process.

The Gamma process is a continuous-time stochastic process, leveraging the characteristics of the Gamma distribution to model positive, variable, and asymmetrically distributed increments. In contract to ex. the Poisson process it can detail the magnitude of efficiency improvements, rather than just counting occurrences. Furthermore the gamma distribution allows for modelling both the average rate of improvement and the variability around this average ensures that both periods of rapid advancements and plateaus can be appropriately represented (Steutel and Thiemann, 1989).

The Gamma distribution for an increment $G(t)-G(s)$ is given by the probability density function:

$$f(x;\, \alpha, \theta) = \frac{x^{\alpha(t-s)-1} e^{-\frac{x}{\theta}}}{\theta^{\alpha(t-s)} \Gamma(\alpha(t-s))}, \quad x > 0 \qquad \textit{eq. 57}$$

The parameters for the process were calculated using method of moments by equating the sample moments (mean and variance) to the theoretical moments of the Gamma distribution. The goodness-of-fit of this appoach was tested applying the KS test on the increments using a 5% significance level. The results did not provide any statistical basis to reject the null hypothesis asserting that the observed data follows a Gamma distribution. 10000 scenarios were simulated using a dedicated *stochastic* library.

Hence the maximum rated power output for a new module each timestep $P_{max}(t)$ is calculated as:

$$P_{PV,new}(t) = P_{PV,new}(t=0) \times (Eff(t) + 1) \qquad \textit{eq. 58}$$

where:

$Eff(t)$ is the new cell efficency at timestep t

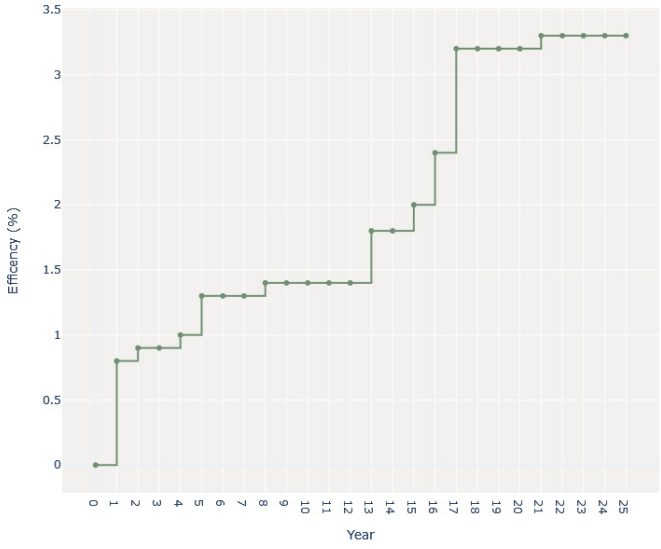


Figure 46: Sample Gamma Process for the PV Cell Improvement

Consequently, under the assumptions that the proportion of losses attributed to inverters remains constant, the power output of a new PV panel each timestep $P_{PV,new}(t)$ is equal to:

$$P_{max}(t) = P_{max}(t=0) \times (Eff(t)+1) \quad \textit{eq. 59}$$

5.7 PV CAPEX

CAPEX for the photovoltaic module was going to be modelled as GBM with a Poisson process (section 3.2.). Historical data regarding the cost eur/Wp of the photovoltaic tehnology was collected for the european market (Appendix J). The analytic solution for this process is given as:

$$Panel(t) = Panel(0)\exp\left[\left(\mu - \frac{\sigma^2}{2} - \lambda_{PT}\theta\right)t + \sigma B_t\right] \times \prod_{i=1}^{N_t}(Y_i) \quad \textit{eq. 60}$$

where:
$Panel0$ is the panel price at the start of the process,
μ is a constant drift rate coefficient,
σ is a volatility coefficient of the process,
λ_{PT} is a mean number of arrivals per unit time,
N_t is a Poisson process with a λ_{PT} frequency rate,
$\theta = E[Y_i\text{-}1]$ where $[Y_i\text{-}1]$ is the random variable percentage change in the cost if the Poisson event occurs.
Y_i corresponds to independent jumps.

An effort was made to calibrate parameters of the jump-diffusion process using maximum likelihood estimation, where the maximum log-likelihood was calculated numerically through Mean-Variance Mapping Optimization. However, this method proved to be problematic, and due to time constraints of the project, an alternative approach was subsequently pursued.

Hence the methodology described in 2.4.5 regarding a GBM with a Learning Rate was used (eq. 25). The learning rate parameter β is assumed as 0.322 (Jung et al., 2021).

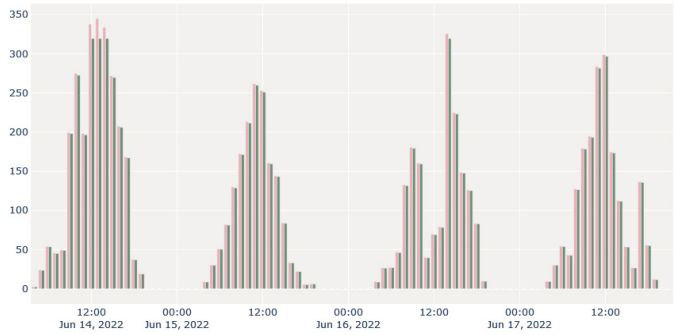


Figure 47: Hourly PV production AC and DC

BoS costs ($BoS(t)$)	50% of the panel costs if new installation, 25% if replacement	IRENA, 2022
Instalation costs ($Inst(t)$)	150 € for the first panel, if more than one panel is installed it decreases by a factor of by 0.1 for each installed module per timestep.	-
OPEX ($OPEX(t)$)	(40 €/kW) x $i_R^{\,t}$	Norouzi et al., 2023
Re-sale value ($Resale_{xi}(t)$)	decreases 10% each year since installation	Price, 2024

Table 12: Cost Variables

TOU (peak)	140% of the base rate	-
TOU (off-peak)	85% of the base rate	-
Feed-in Tariff	1/3 of the electricity tariff	Norouzi et al., 2023

Table 13: Electricity tariff rates

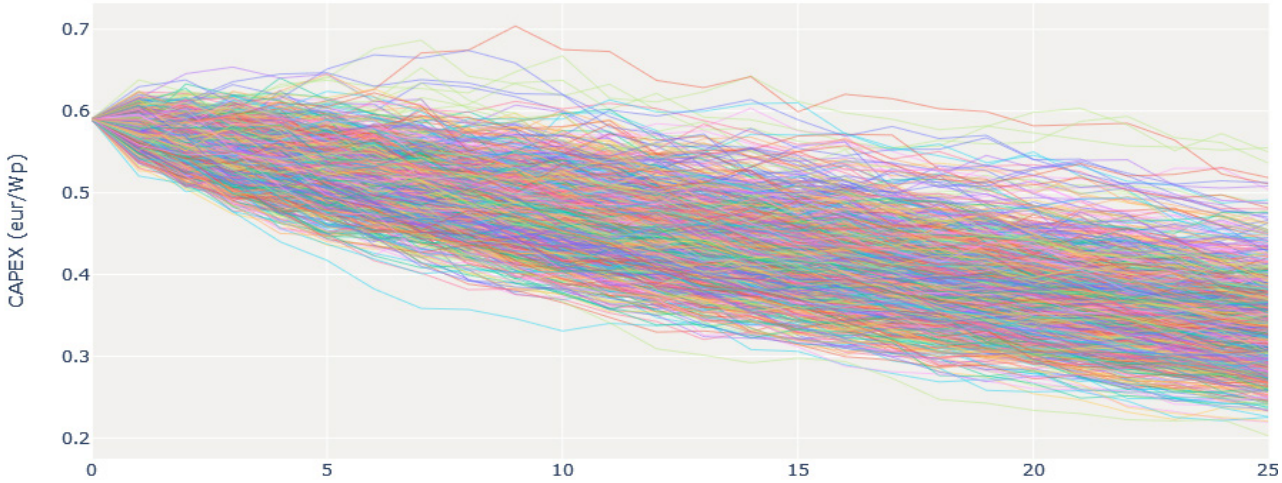


Figure 48: The generated scenarios of the CAPEX for JA module + Inverter price

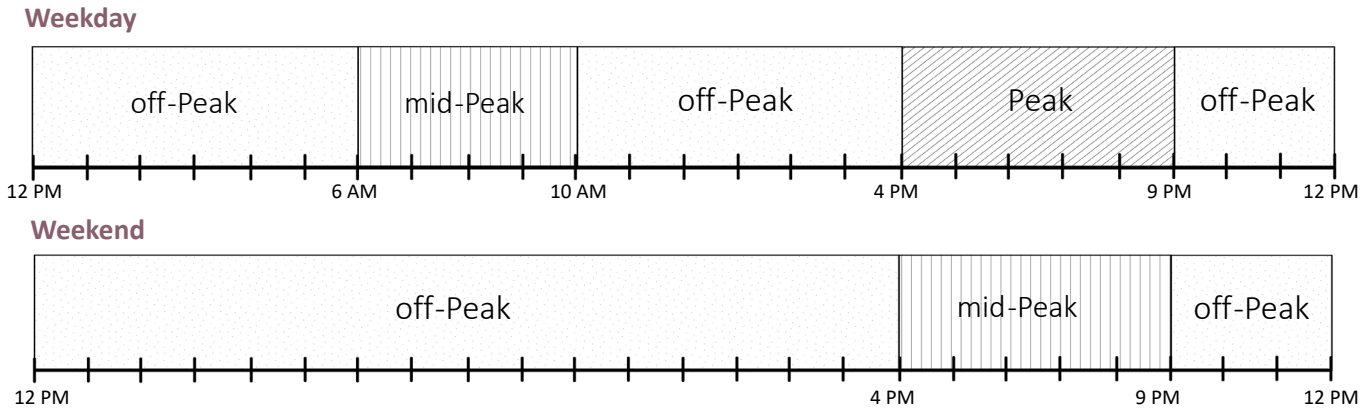


Figure 49: Time-of-Use Electricity Rates

Variable	Value
Inflation rate (i_R)	0.025
Loan Interest rate (i_L)	0.06
Discount factor (ϑ)	0.04

Table 14: Financial Rates

The future price of PV modules in the Netherlands was estimated using projected cumulative solar installation capacity data from, Vartiainen et al., 2019 (base scenario) which then facilitated the calculation of the mean drift. The volatility was on the other hand calibrated from the past data (Appendix J).

10000 scenarios were generated for each base panel (fig. 48). Subsequently the price of the module is adjusted to the PV cell efficency increase over time and hence the increase of the maximum rated power output $P_{max}(t)$ in Wp:

$$Price(t) = P_{max}(t) \times CAPEX(t) \qquad eq.\ 61$$

where:
$CAPEX$ is the capital costs eur/Wp at timestep t
$Price(t)$ is price for panel + inverter in eur at timestep t

5.8 Time-of-Use

The Time-of-Use pricing scheme is assumed for the ectricity tariff, described in the section 2.2.2.3. The cost of electricity is based on the time of day, reflecting the fluctuating demand and supply conditions in the power market. Under this tariff structure, electricity prices are divided into different rates for specific periods: peak, off-peak, and mid-peak hours. Those rates are visualised on figure 49 and given in table 13.

5.9 Electricity Balance

The electricity balance within the system must adhere to the following constraints:

$$P_{load}(h) = P_{PV,total}(h) + E_{sup}(h) \qquad eq.\ 62$$

$$P_{PV,total}(h) = P_{load}(h) + E_{exp}(h) \qquad eq.\ 63$$

$$E_{exp}(h) \leq E_{exp,max} \qquad eq.\ 64$$

$$E_{exp}(h) \times E_{imp}(h) = 0 \qquad eq.\ 65$$

where h is an hourly timestep.

The electricity balance within the system must achieve equilibrium. The electricity demand must be met by combining the electricity imported from the grid and the yield from photovoltaic modules. Concurrently, the electricity generated by the PV panels is either consumed locally or exported back to the grid, provided it does not exceed the maximum export wattage ($E_{exp,max}$ = 4kW). Electricity cannot be simultaneously sold to and purchased from the grid.

5.10 PV Module Carbon Intensity

To determine the values for the kg of CO2 equivalent embodied in the PV installation, including the module, balance of system (BoS), transport, and inverter, an extensive literature review was conducted. This review highlighted that data on the carbon footprint from life cycle assessments (LCA) of PV modules for on-roof use in Europe is sparse and highly variable (section 2.3.2.2). Consequently, it is not feasible to simulate this stochastically due to the lack of reliable historical data.

The total GHG emissions for timestep 1 were calculated based on the findings of Krebs, Frischknecht, and Stolz (2020). Their study determined the carbon intensity of a monocrystalline PV array, including the inverter and mounting equipment. From their research, it was calculated that the carbon footprint (CO2) for 1 kWp is 1620 kg CO2eq/kWp. This value was applied to both the base and budget models, while the second-hand models

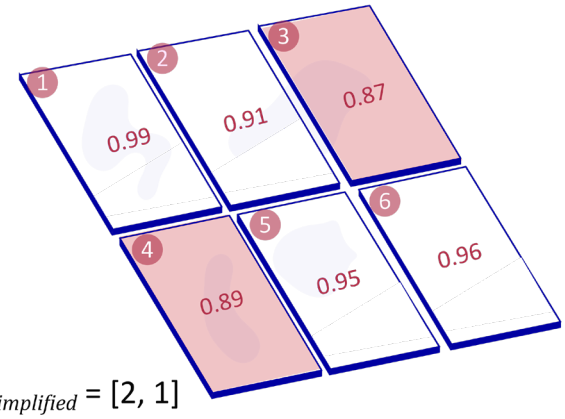
```
to_replace = action[0:number_of_panels-1].sum()
panel_type = action[number_of_panels]

# Find indices of the lowest 'action' values in previous_observation
indices = []

for count, i in enumerate(action[:self.number_of_panels]):
    if i == 1:
        indices.append(count)

# Replace these indices in the observation with efficiency_develop
self.observation[:self.number_of_panels][indices] = self.efficency_develop


# Copy over the other values from previous_observation to observation
mask = np.ones(self.number_of_panels, dtype=bool)
mask[indices] = False
self.observation[:self.number_of_panels][mask] = self.previous_observation[:self.number_of_panels][mask]


# Mark panels that were replaced
replaced_panels = np.zeros(self.number_of_panels, dtype=int)
replaced_panels[indices] = 1

# Set to the corresponding panel ID based on action[1]
for index in indices:
    self.observation[-self.number_of_panels + index] = self.panel_ids[panel_type]
```

Figure 50: Observation update in the RL enviornmnent, under the complex action space

were assumed to have no impact.

For the timestep at t=25, the findings of Frischknecht, Itten, and Wyss (2015) were used. Following the same methodology as described above, the carbon emissions (CO2) under the REAL scenario, which assumes realistic improvements until 2050, were estimated to be 833 kg CO2eq/kWp.

Both studies specifically refer to on-roof residential PV installations under Central European conditions. The values from these studies were linearly interpolated to account for all values between the start and the end of the optimization period. It was decided not to include the carbon footprint in the observation space due to its high uncertainty and the very possible lack of recent relevant data when implemented in practice.



$A_{simplified}$ = [2, 1]

$A_{complex}$ = [0, 0, 1, 1, 0, 0, 1]

Figure 51: Comparison of complex and simplifies action space

The carbon emissions per each installed panel are given as:

$$CO_{2,PANEL} = CO_2(t) \times P_{max,PANEL}(t) \qquad eq.\ 66$$

where:
$CO_{(2,PANEL)}$ is the installed panel's total embodied carbon,
$P_{max}(t)$ is the installed panel's power capacity in kWp,
$CO_2(t)$ are the mean carbon emissions per 1kWp.

5.11 Action Space

The exploration of the action space has been conducted with increasing complexity of the environment in mind. Initially, the plan was to utilize for this phase an action space mirroring the one used in the toy problem. However, during the testing phase, it was observed that this approach caused the agent to struggle in achieving optimal convergence and also resulted in slower performance, hindering the efficiency of the agent in reaching its objectives. Hence, an alternative **simplified discrete action space** was developed facilitating faster and much more stable convergence in comparison.

Under the simplified action space in the single-panel scenario, it is represented by a single integer that ranges from 0 to the maximum number of grid cells, indicating the number of panels exchanged in the current timestep. The option to discart a panel is removed. Panels are

chosen for exchange based on their efficiency, starting with the least efficient.

In the multi-panel scenario, the simplified action space is multidiscrete and comprises two integers: the number of panels installed and the type of panel (table 9). It is assumed that only one module type can be selected for exchange each year. This can be expressed as:

$$A = \{(x, y) \mid 0 \leq x \leq N \; and \; 0 \leq y < M\} \qquad eq.\ 67$$

where:
x is the number of panels exchanged,
y is the type of module chosen,
N is the maximum number of panels,
M represents the maximum number of panel types.

Alternatively if there are multiple roof surfaces (as in house 3) with varying irradiance and shading, x is expressed in more integers, each one for each surface ex. *x1, x2* for house 3 (double-pitched roof).

On the contrary the **complex** fully multi-discrete **action space** while training cosidereably slower, theoretically, its complexity should allow for more refined decision-making, potentially leading to superior outcomes, particularly under conditions not examined in this study, such as varying shading or mounting configurations of panels within a single system. Similarly to the simplified action space for the multi-panel scenario only one panel type can be chosen each timestep. It is expressed as:

$$A = \{(x_i, x_{i+1} \ldots, x_n, y) | 0 \leq x_i \leq 1 \; and \; 0 \leq y \leq M\} \qquad eq.\ 68$$

where n is the number of panels.

Consequently, it was determined that both action spaces should be evaluated to assess their relative effectiveness. Upon receiving the action A from the agent, the environment checks whether the current time step t is not the final one, t≠H. If this condition holds, it proceeds with the following updates (figure 50). Both action spaces are visualised on figure 51. Here both actions result in the same outcome of replacing panels 3 and 4.

Budget (at t = 1)	**Value**
Low	0 - 750 eur
High	750 - 2000 eur

Table 15: Budget ranges at t=1

5.12 Observation Space and Reward

The observation space is multi-continuous, comprising a tuple of integers for each relevant observation. These observations include integers for the installation of the panels, their current efficiency ranging from 0 to 1, the current average electricity tariff, the feed-in tariff (export price), the current market efficiency of PV modules, the CAPEX of the new module, the carbon emissions from the grid (for the variant with an environmental aspect), the current budget, previous budget, previous expenditure and finally a panel *id* for every installed panel, for the multi-panel scenario.

$$OBS(t) = [\eta_i(t), \eta_{i+1}(t), \ldots \eta_{i+n}(t), I(t), J(t), Eff(t), \; CAPEX_{PV}(t), C(t-1), Pen(t), B(t), B(t-1), EG(t), id_i(t), id_{i+1}(t), \ldots id_{i+n}(t)] \qquad eq.\ 69$$

Similarly to the toy problem the financial balance is defined as:

$$Balance_{FIN}(t) = \{[-C(t) - I(t) \times E_{\sup}(t) + J(t) \times E_{\exp}(t) + Resale_{tot}(t)] - Pen(t)\} \times (1 - \vartheta)^t \qquad eq.\ 70$$

where:
$I(t)$ represents is electricity tariff at year t,
$J(t)$ represent the compensation tariff at year t,
ϑ is the discount factor, set to 1
$C(t)$ the installation costs is calculated as:

$$C(t) = CAPEX(t) + OPEX(t) + BoS(t) + Inst(t) \qquad eq.\ 71$$

$Pen(t)$ is the penalty for exceeding the maximum annual budget.
$Resale_{tot}(t)$ is the value of operational panels sold each step.

The approximations and assumptions for the Balance of System, Installation, and Operational Expenses are indicated in table 12. Capital expenditures, BoS, and Installation costs are multiplied by the number of panels added in the given timestep. Operational expenses are calculated based on the total kW of currently installed panels.

$Pen(t)$ is the annual repaynent of the loan interest for exceeding the budget $B(t)$ (table 15). Annual budget is adjusted for inflation. The loan terms for a given budget breach are elaborated on in Appendix H. Penalty for each timestep is shown in equation 72 where:

$$Pen(t) = \sum_{i=1}^{t-1} \left(\max\left(C(i) - B(i), 0\right) - \sum_{j=i=1}^{t} \frac{\max\left(C(i) - B(i), 0\right)}{LoanTerm(i)} \right) \times {i_L}^i \times 1_{\{(t-1) \leq LoanTerm(i)\}} \qquad eq.\ 72$$

LoanTerm as the duration of the loan,
LoanAmount(t) as the amount of loan taken at time t,
LoanAmount(t) = max(C(t)−B(t), 0),
Repayment(t) as the amount repaid at time t,
The Reward for the scenarios with exclusively focusing on financial optimisation is:

$$R_{FIN} = \sum_{t=1}^{25} BalanceFIN(t) \quad \textit{eq. 73}$$

On the other hand the environmental balance is:

$$Balance_{ENV}(t) = CO_{2,PANEL}(t) - \left[EG(t) \times \left(E_{PV,total}(t) - E_{PV,dumped}(t)\right)\right] \quad \textit{eq. 74}$$

where:
$\sum CO_{2,PANEL}(t)$ is the cumulative installed module's embodied carbon,
$EG(t)$ is the grid factor,

$$E_{PV,dumped}(t) = \sum_{8760}^{h} E_{exp}(h) - E_{exp,max} \quad \textit{eq. 75}$$

Hence the reward for the optimisation variant with both the financial and the environmental factors considered is:

$$R_{tot} = \sum_{t=1}^{25} Balance_{FIN}(t) \left[w + \sum_{t=1}^{25} Balance_{ENV}(t)\right] \quad \textit{eq. 76}$$

The distribution of both balances is normalised to identical means and standard deviations.

5.13 Evaluation Metrics

In contrast to the toy problem where the internal rate of return (IRR) was employed for financial evaluation, this chapter opts for net present value (NPV) as the preferred metric. The rationale for this selection stems from the limitations of IRR, which is applicable primarily when expenditures (negative cash flows) occur exclusively in the initial year. However, in the current model, the agent's actions may lead to negative cash flows in subsequent steps, thereby necessitating a metric that can accommodate such financial dynamics throughout the entire period under consideration. Thus, NPV is chosen for its suitability in scenarios where cash flows vary over time. The balance considered for calculating the NPV is almost identical to the one used for the reward function, with the exception that instead of accounting for the costs of electricity imported from the grid, it considers the electricity produced by the panels that is directly consumed by the household load. This approach focuses on the electricity savings resulting from the self consumption.

$$Balance_{NPV}(t) = \{[-C(t) + I(t) \times E_{used}(t) + J(t) + E_{used}(t) + Resale_{tot}(t)] - Pen(t)\} \times (1 - \vartheta)^2$$

where: *eq. 77*

$$E_{used} = \min\left(E_{PV,total}, E_{load}(t)\right) \quad \textit{eq. 78}$$

So the NPV is:

$$NPV = \sum_{t=0}^{T} \frac{Balance_{NPV}(t)}{(1+\vartheta)^t} \quad \textit{eq. 79}$$

The environmental benefit on the other hand is evaluated as Net Carbon Savings (NCS). This metric takes into account the emissions associated with the lifecycle of the PV array itself and subtracts the carbon emissions avoided by using PV-generated electricity instead of grid electricity:

$$NCS = \sum_{t=1}^{T} Balance_{ENV}(t) \quad \textit{eq. 80}$$

Model results are compared with the **base policy**, where panels are replaced once they fail or reach 80% of their original efficency, which is based on the common practice in the GCRS.

For the select panel scenario the type of the panel is determined by calculating the expected NPV for each type considering the current electricity tariff, mean degradation rate and expected lifespan (φ, table 8).Due to time constraints, it was decided to focus training and evaluation of the complex action space exclusively on scenarios involving multiple panels, as these were deemed more practically significant.

5.14 Model Selection and Parameters Tuning

Before training all datasets were split in proportion 7:3 for training and evaluation respectively. A seperate evaluation environment was built that included the evaluation metrics described in the previous section.

A reward normalisation was attepted by scaling reward values using an incremental algorithm for maintaining a running average and variance, to ensure that the scale of rewards does not affect the learning updates disproportionately. However it did not have a significant effect on the training.

5.13.1 Agent Selection

A study was conducted to select the optimal agent

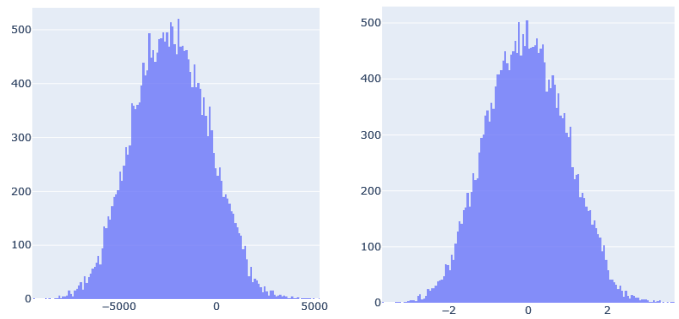


Figure 52: Reward Distribution under random actions - not normalised and normalised

architecture. In this study, similarly to the toy problem, the performances of the A2C and PPO algorithms were evaluated under identical training parameters (table 16). Different subprocessing strategies were tested.

PPO generally exhibits more stable performance compared to A2C, particularly in environments with the complex action space. The number of environments seems to affect the algorithms differently. PPO's performance improves slightly with more environments, while A2C shows significant volatility and generally lower performance as the number of environments decreases. The complexity of the action space has a profound impact on A2C's performance, leading to greater instability and lower rewards. Overall, PPO appears more robust across different numbers of environments and varying complexities in the action space, making it more suitable choice for the task.

5.14.3 Hyperparameter Tuinng with Optuna

After the selection of the algorythm and the policy an

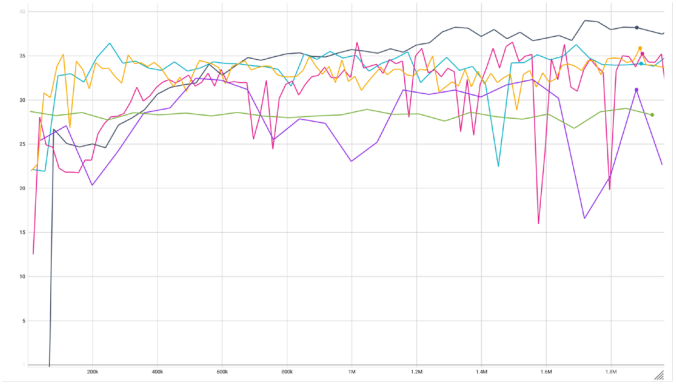


Figure 53: A) Evaluation Reward - PPOs: 16 envs - black, 8 envs - pink, 4 envs - purple; A2Cs: 16 envs - light blue, 8 envs - yellow, 4 envs - green (simple action space)

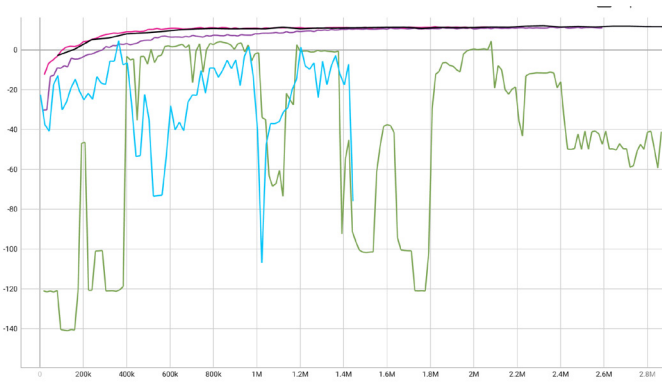


Figure 43: B) Evaluation Reward - PPOs: 16 envs - black, 8 envs - pink, 4 envs - purple; A2Cs: 16 envs - light blue, 4 envs - green (complex action space)

automatic hyperparameter tuning was conducted using the python Optuna library. Optuna uses a tree-structured parzen estimator as the sampler for the hyperparameters selection by modeling the probability distribution

Neural Network	64, 64
Activation Function	Tanh
Learning Rate	0.0003
No of Environments	16
Gamma	0.99
Lambda	0.95
Batch Size	64
Clip Ratio	0.2
Entropy coefficient	0.0
No of Epochs	8

Table 16: Deafult Parameters

No of Neurons	64, 128, 256
No of Layers	2, 3
Activation Function	Tanh, ReLU
Learning Rate	0.01 - 0.00001
No of Steps	1024 x 1, 2, 4, 8
Gamma	0.95 - 0.99
Lambda	0.90 - 0.95
Batch Size	64, 128, 256
Clip Ratio	0.15 - 0.25
Entropy coefficient	0.0 - 0.02
No of Epochs	8 - 32

Table 17: Ranges of the parameters for optuna optimisation

No of Neurons	256
No of Layers	2
Activation Function	Tanh
Learning Rate	0.00026
No of Steps	2048
Gamma	0.99
Lambda	0.93
Batch Size	256
Clip Ratio	0.17
Entropy coefficient	0.0
No of Epochs	24

Table 18: Baseline Model Parameters

of hyperparameters given the scores of past trials.

Additionally, a median pruner is used that stop the evaluation of a trial if it is judged unlikely to result in a competitive outcome, in order to save computational power and time.

The hyperparameters targeted for optimization were specified in various formats: as discrete uniform distributions, simple categorical lists, or as ranges of floating-point numbers. Details regarding these hyperparameters, their potential ranges, and possible values are provided in Table 17.

A study conducted involved a total of 150 trials, each running for one million timesteps. A pruning mechanism was implemented, terminating any trial that underperformed after 200,000 timesteps. Figure 46 displays partial results, illustrating the performance of each hyperparameter against the achieved scores.

Overall the NN architecture that yeilded the highest score Overall, the neural network architecture that yielded the highest score consisted of 2 layers, each with 256 neurons. Altering the activation function did not significantly impact the results. The optimal learning rate was approximately 0.0003, which aligns with the default value used in stable baselines. The best results were

No of Neurons	256, 512, 1024, 2048
Batch Size	256, 512, 1024, 2048
Learning Rate Schedule	Flat, Linear, Exponential
Target Learning Rate	0.0001 - 0.00001

Table 19: Ranges of the hyperparameters for the manual optmisation

achieved by setting the batch size and discount factor (γ) to their upper limits, while reducing the number of steps and the entropy coefficient to their lower bounds.

The parameters from the trial that yielded the best results have been selected for further evaluation as the Baseline Model (table 18).

5.14.4 Manual Tuning

After establishing the Baseline Model, two hyperparameters—minibatch size and number of neurons—were identified as crucial for achieving optimal scores based on the results. Both parameters were set to their maximum values in the Optuna optimisation, suggesting that further increases might lead to improved

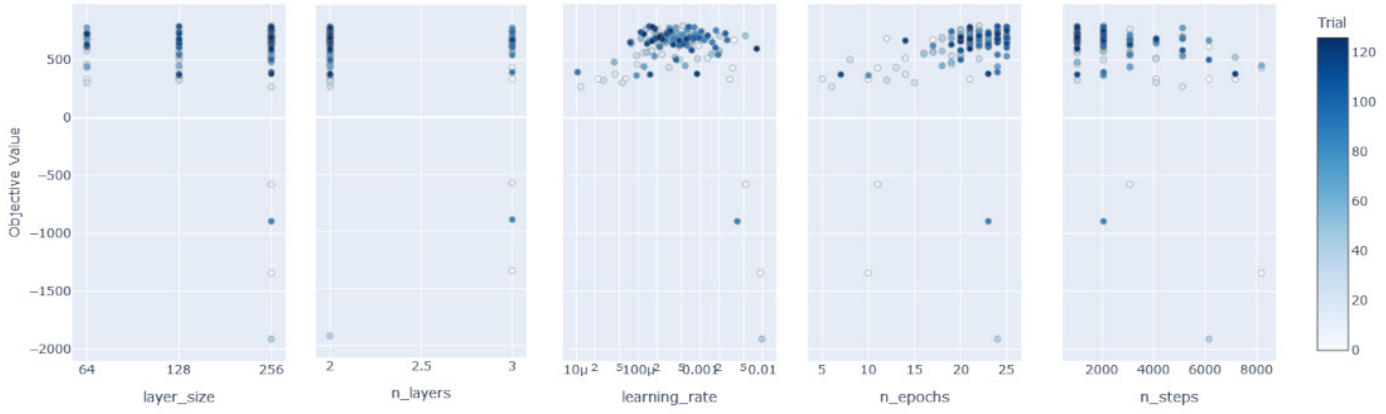


Figure 60: Comprahensive Plot of the NPV results for each

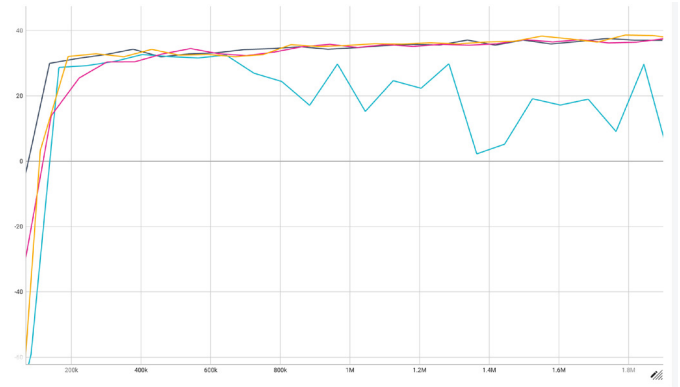


Figure 54: Evaluation Reward for different network sizes for 512 batch size: 256 - blue, 512 - pink, 1024 - blue, 2048 - black (Simple Action Space)

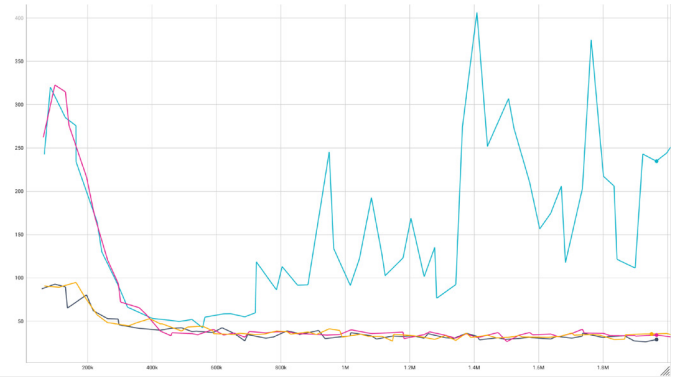


Figure 55: Total loss for different network sizes for 512 batch size: 256 - blue, 512 - pink, 1024 - yellow, 2048 - black (Simple Action Space)

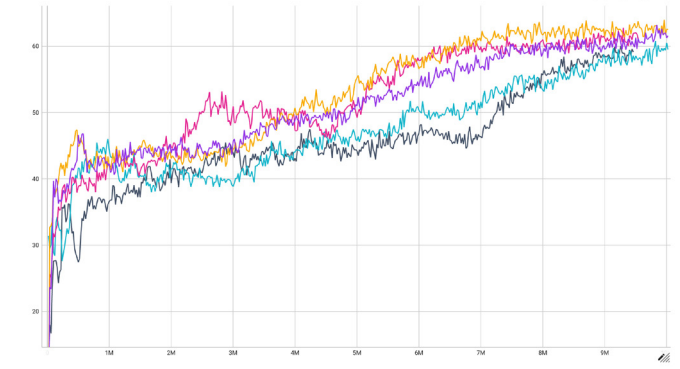


Figure 56: Evaluation Reward for different mini batch sizes (1024, 1024 NN): 256 - black, 512 - pink, 1024 - yellow, 2048 - blue (Complex Action Space)

outcomes. Additional experiments were conducted by adjusting the learning rate through the implementation of a learning rate scheduler, which decreases the rate as training progresses (the learning rate at the start is always 0.0003).

Regarding the slow speed of the optuna study it was decided to perform this part manually.

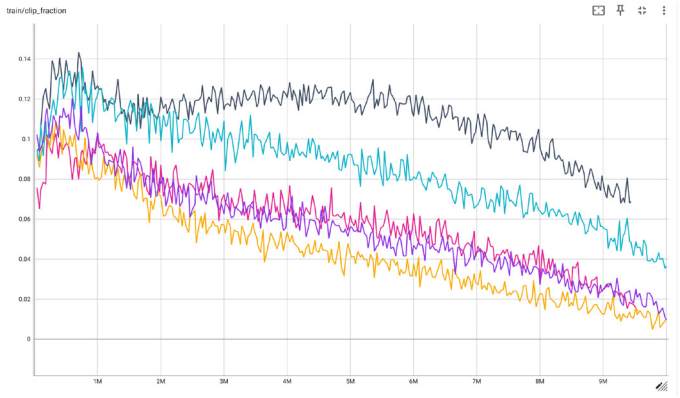


Figure 57: Clip Ratio (fraction of time the policy's probability ratio is clipped) for different mini batch sizes (1024, 1024 NN): 256 - black, 512 - pink, 1024 - yellow, 2048 - blue (Complex Action Space)

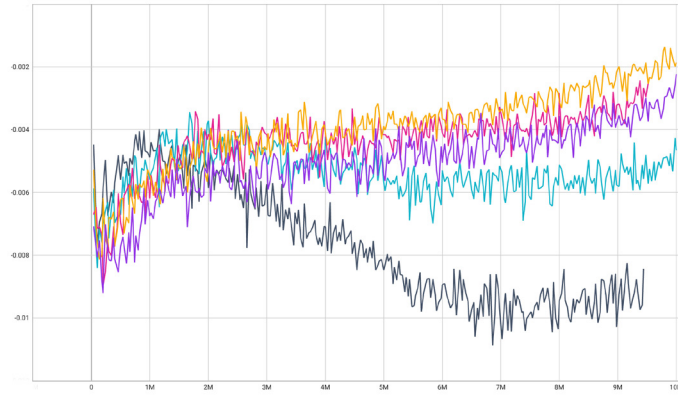


Figure 58: Policy Gradient Loss for different mini batch sizes (1024, 1024 NN): 256 - black, 512 - pink, 1024 - yellow, 2048 - blue (Complex Action Space)

The hyperparameters considered in this phase are outlined in table 19.

In total 25 different models containing different combinations of those hyperparameters were trained. Each one was then evaluated against the Baseline Model both on two distinct environments: budget panel, low budget, house 2 and select panel, high budget, house 1 with the complex action space for scalability. Those were evaluated on the basis of their evaluation reward and training metrics such as approximate Kullback-Leibler divergence, explained variance, training loss and policy and value losses. The results of this phase are shown in Apppendix L.

The parameters of the final model are shown in the table 20. All models are trained on HP Z-Book Studio G7 using an Intel Core i9-10885H CPU @ 2.40GHz. It was attempted to train using its NVidia Quatro T2000, however the model fitting was slower.

No of Neurons	1024
No of Layers	2
Activation Function	Tanh
Learning Rate	Linear decay from 0.0003 to 0.00001
No of Steps	2048
Gamma	0.99
Lambda	0.93
Batch Size	1024
Clip Ratio	0.17
Entropy coefficient	0.0
No of Epochs	24

Table 20: Final Model Parameters

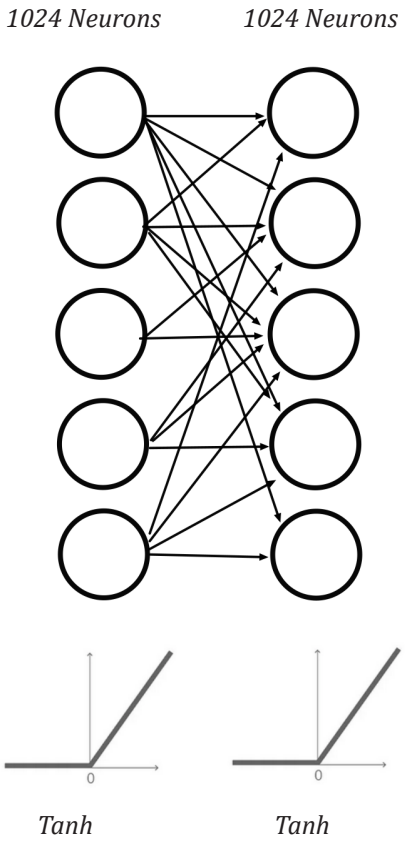


Figure 59: Final NN for both the actor and critic

5.15 Results

The outcomes of the model evaluation, specifically the Net Present Values and Net Carbon Savings are detailed in Tables 21 , 22 and 23 for houses 1, 2 and 3, where they are compared against the base policy. The plot of NPVs for each analysed scenario under financial reward is shown on figure 61. Due to time constraints the training and evaluation of the reward function including both the financial and environmental balances was conducted for the house 3 only. Evaluation results for House 3 are depicted on figures 62 and 63.

The NPV distributions for the scenario involving a budget panel with a high budget for house 2 are depicted in Figure 65. Additionally, a detailed plot for a selected episode within this scenario is illustrated in Figure 66. The results for each scenario are discussed below.

5.15.1 House One

The scenario involving the second-hand panel consistently shows the trained model underperforming compared to the base policy. This underperformance is primarily attributed to the low capital expenditure (CAPEX) of the modules, which consequently does not lead to budget overruns, particularly in the high-budget variant. In this variant, the average interest for an episode under the base policy amounts to 62 EUR, representing 0.027% of the total expenditure for that episode. In contrast, the average interest for the model evaluation is zero, indicating that the agent strictly avoids exceeding the budget. As a result, this leads to a more conservative strategy regarding the replacement of degraded panels, with the model replacing panels at a mean rate of 76% compared to the base policy's 80%.

In the case of the budget panel, the model yielded suboptimal results compared to the base policy in both budget scenarios. The agent chose to install all possible panels in the first timestep in only 11% of the episodes. While this strategy helped avoid accruing significant interest—averaging €32 in interest after the first step in the agent evaluation, compared to €57 in the base policy—it similarly led to overly conservative behavior. In the low budget scenarios, the total power typically did not exceed 4,500 Wp by timestep 3, particularly when themodule's CAPEX remained steady or increased marginally.

In the high-end panel scenario, the model outperformed the base policy. However, it is important to note that despite yielding the highest energy among all panel types, the Net Present Values for the high-end panels are consistently lower than those for the budget panel. This outcome is attributed primarily to the higher CAPEX associated with the high-end panels and the relatively short optimization period, which is significantly shorter than the average failure rate of these panels. Consequently, there is limited benefit in choosing these more expensive panels over the cheaper alternatives. Nonetheless, the agent's conservative strategy in panel adoption proved successful. Under the base policy, the mean interest accrued after the first step was €187, compared to €76 in the agent evaluation. Overall, the additional interest costs amounted to 20% of the base policy expenses, versus 9% for the model evaluation.

In the select panel scenario, as anticipated, the simplified model outperformed the complex action space and narrowly exceeded the base policy. Overall, the agent using the simplified action space more frequently chose the second-hand panel in the early stages of the optimization process, consistently selecting the budget panel thereafter. In contrast, the complex action space failed to make these selections, instead sporadically opting for the high-end panel, which is financially impractical at this level. The base policy consistently selected the budget panel.

Furthermore, both trained agents exhibited more conservative behaviour than the base policy, typically replacing panels with an average 78% efficiency. The complex action space model, in a vast majority of cases, chose the least efficient panel for replacement. Additionally, in 4% of all evaluation episodes, it failed to replace a panel that had malfunctioned in the previous step, suggesting that the algorithm might be overstretched.

5.15.2 House Two

Not unlike the house 1, the scenario involing the 2nd-hand module is when the model performs the worst then benchmarked against the base policy. Here however, the increased complexity of the environment, the higher maximum number of panels, and consequently the greater potential to exceed the budget or make poorly timed decisions contribute to the model's improved performance compared to the base policy. This improvement is consistent across all scenarios for the simplified action space, as illustrated in Figure 51. The agent effectively maintains a stable output from the PV array throughout the optimization period and strategically replaces panels when CAPEX decreases and the budget permits. This management approach results

		Low Budget			High Budget		
		Base Policy	*Model*		*Base Policy*	*Model*	
Action Space		-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	5 633	-	5 453 (-2.1%)	6 622	-	6 350 (-3.6%)
Budget Panel	*NPV*	9 584	-	9 498 (-0.9%)	11 571	-	11 363 (-1.2%)
Hi-End Panel	*NPV*	7 705	-	7 890 (+2.4%)	8 880	-	9 022 (+1.6)
Select Panel	*NPV*	7 894	7 728 (-5.1%)	8 068 (+2.2%)	8 190	7 927 (-7.2%)	8 272 (+1.0%)

Table 21: A) Evaluation Results for House 1 - NPVs

		Low Budget			High Budget		
		Base Policy	*Action Space*		*Base Policy*	*Action Space*	
Action Space		-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	10 821	-	11 164 (+1.6%)	10 857	-	10 792 (-0.8%)
Budget Panel	*NPV*	18 677	-	20 060 (+7.4%)	19 192	-	19 598 (+2.1%)
Hi-End Panel	*NPV*	16 023	-	17 193 (+7.3%)	17 098	-	18 054 (+5.6)
Select Panel	*NPV*	16 595	15 936 (-18%)	17 508 (+5.5%)	17 413	16 497 (-15%)	17 988 (+3.3%)

Table 22: Evaluation Results for House 2 - NPVs

in a smoother distribution of interest costs over the optimization period, in contrast to the base policy which exhibits significant jumps.

Moreover, in scenarios with a high-end budget, the agent prudently refrains from installing all available panels until the CAPEX drops to around 70% of its initial value.

The distributions of NPVs for the base policy and the model with simplified action space were analyzed, revealing distinct differences in performance and variability. Figure 65 (A shows the bar distribution for the model's evaluation, which predominantly clusters around 30K with a narrower spread and fewer outliers, indicating a more consistent performance. Conversely, Figure 65 (B illustrates the base policy's NPV distribution, which is wider, suggesting greater variability in outcomes with a significant tail extending into higher NPVs. The box plots in Figures 65 (C and 65 (D further highlight these contrasts; the model's median NPV is higher and displays a tighter interquartile range, whereas the base policy, although exhibiting a lower median, also shows greater spread and variability.

House Two further demonstrates that the complex action-space-based model is overextended, resulting in the agent's failure to achieve optimal convergence due to its excessive scale (8,388,608 possible action configurations per step, compared to 72 for the simplified action space). The percentage of episodes in which a failed panel is not replaced in the subsequent step increased to 32%. Moreover, the selection of the panel type for installation appears to occur randomly.

		Low Budget	High Budget
		Base Policy	*Base Policy*
Action Space		-	-
Second-Hand Panel	*NPV*	12 813	13 987
	NCS	44 404	
Budget Panel	*NPV*	22 132	29 110
	NCS	39 295	
Hi-End Panel	*NPV*	22 229	24 020
	NCS	45 442	
Select Panel	*NPV*	21 075	24 545
	NCS	38 679	

Table 23: A) Evaluation Results for House 1 - Base Policy

		Low Budget		High Budget	
		Action Space		*Action Space*	
Action Space		*Complex*	*Simplified*	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	-	13 188 (+2.9%)	-	14 205 (+1.6%)
	NCS	-	46 737 (+5.0%)	-	47 517 (+7.0%)
Budget Panel	*NPV*	-	22 241 (+0.1%)	-	29 351 (+0.8%)
	NCS	-	40 028 (+1.9%)	-	36 326 (-8.4%)
Hi-End Panel	*NPV*	-	23 785 (+6.6%)	-	25 959 (+7.5%)
	NCS	-	48 051 (+5.5%)	-	43 501 (-4.4%)
Select Panel	*NPV*	13 543 (-55.6%)	22 461 (+6.2%)	15 616 (-57.2%)	27 065 (+9.3%)
	NCS	31 660 (-22.2%)	41 833 (+7.4%)	31 660 (-22.2%)	37 486 (-3.1%)

Table 23: B) Evaluation Results for House 3 - Financial Balance Only based Reward

		Low Budget		High Budget	
		Action Space		*Action Space*	
Action Space		*Complex*	*Simplified*	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	-	11 488 (-11.5%)	-	11 780 (-8.7%)
	NCS	-	48 003 (+7.5%)	-	48 766 (+9.8%)
Budget Panel	*NPV*	-	21 743 (+1.7%)	-	28 832 (-0.9%)
	NCS	-	41 540 (+5.5%)	-	43 735 (+11.3%)
Hi-End Panel	*NPV*	-	22 252 (+0.2%)	-	24 653 (+2.6%)
	NCS	-	48 946 (+7.2%)	-	46 427 (+2.7%)
Select Panel	*NPV*	11 984 (-56.7%)	22 361 (+5.8%)	14 485 (-59.0%)	25 251 (+2.9%)
	NCS	33 330 (-16.0%)	45 798 (+18.4%)	35 020 (-10.4%)	46 798 (+21.0%)

Table 23: C) Evaluation Results for House 3 - Financial + Environmental Balance based Reward

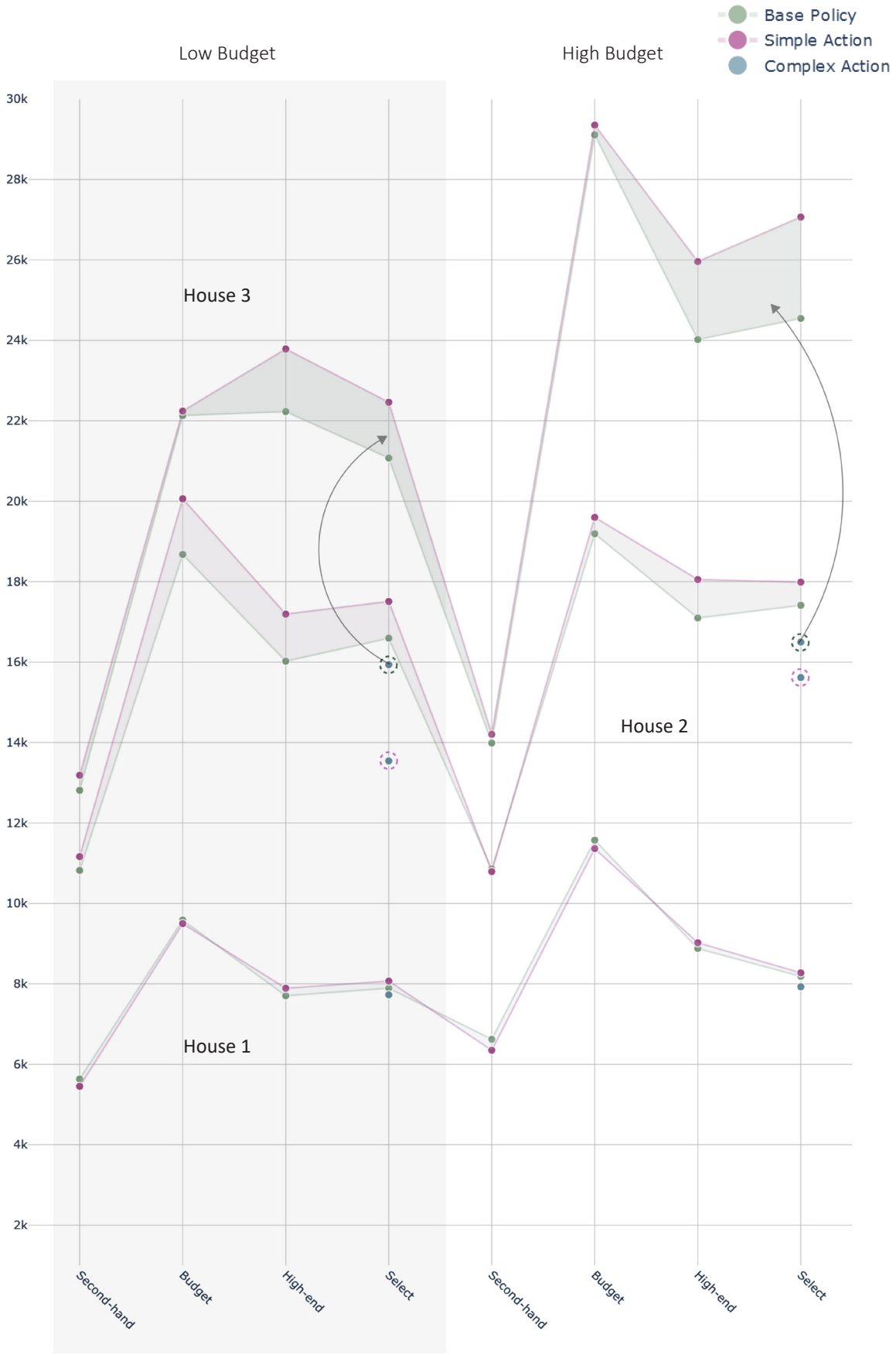


Figure 61: Comprahensive Plot of the NPV results for each analysed scenario under financial balance reward only

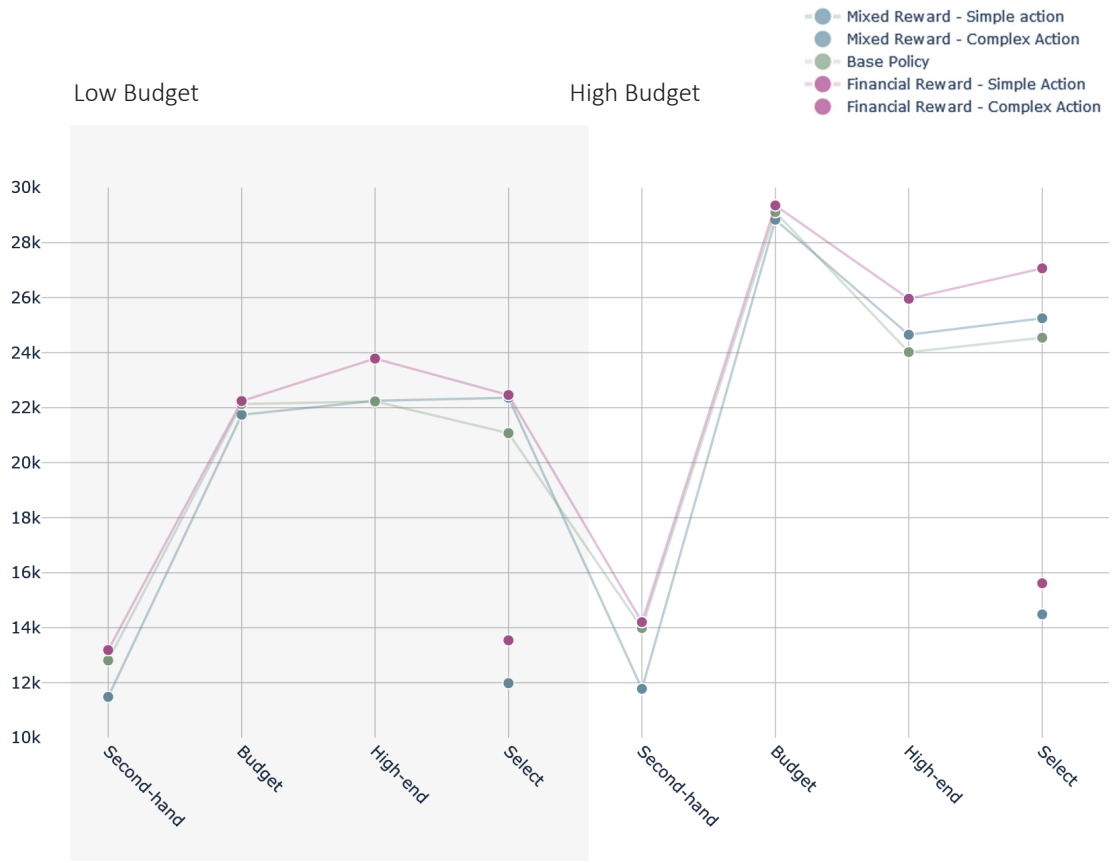


Figure 62: Plot of NPVs for each analysed scenario for House 3

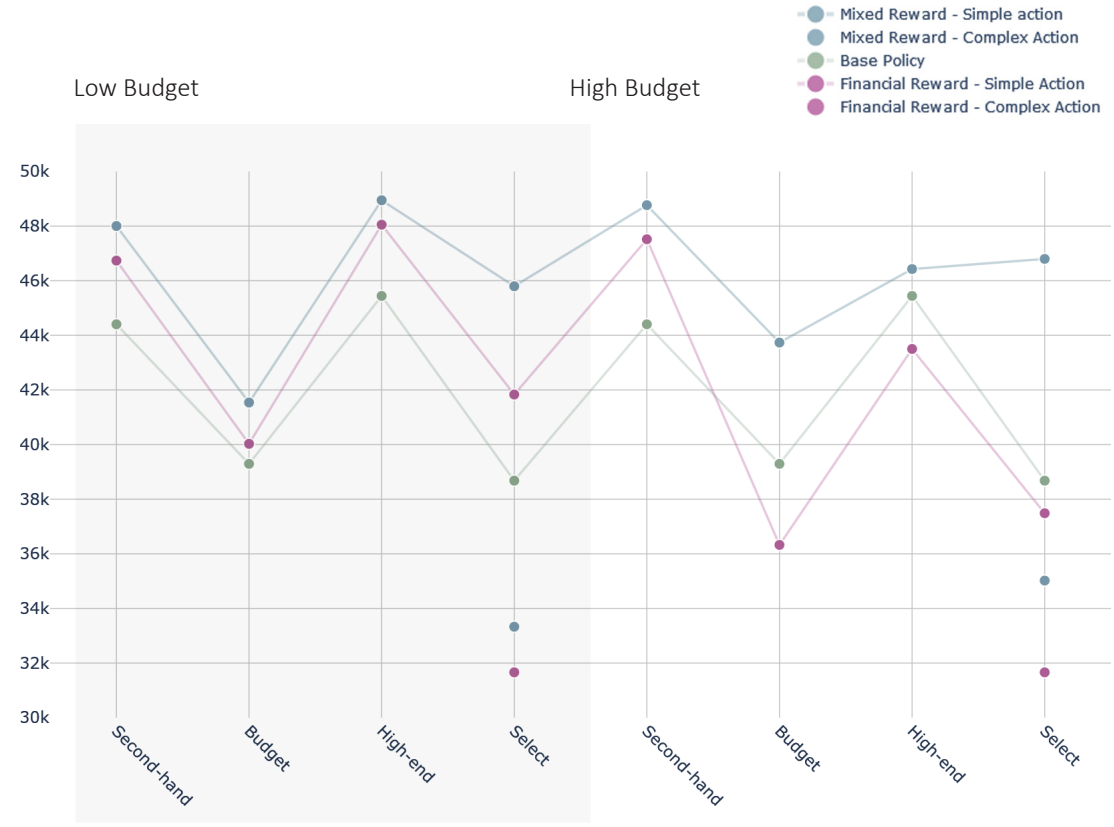


Figure 63: Plot of NCSs for each analysed scenario for House 3

5.15.3 House Three

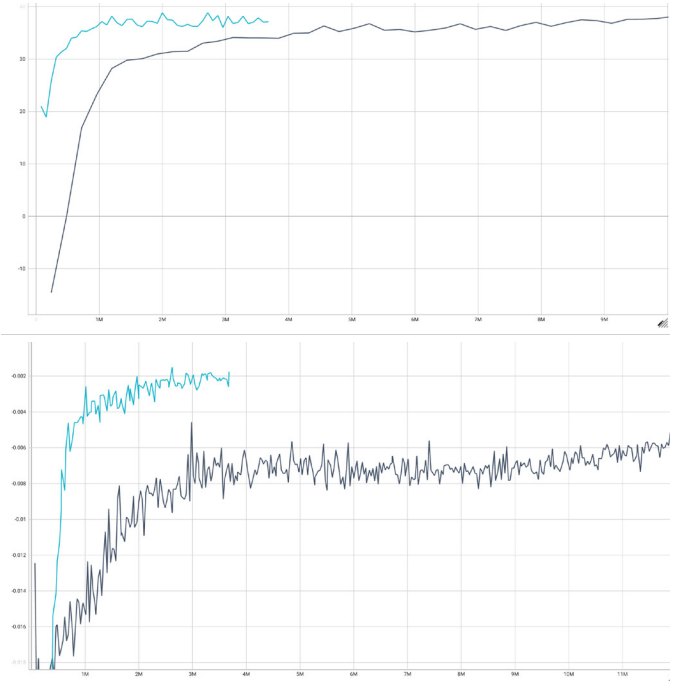
For house three both financial and environmental evaluation was performed, with two reward functions tested.

Financial Reward

When evaluating NPV, the model shows similar patterns to houses 1 and 2, performing significantly better in scenarios with tighter budgets by avoiding extra interest costs. The generally more conservative behavior of the agent, as described in previous sections, benefits the system's carbon savings in lower-budget scenarios. In these cases, the agent adopts a more conservative installation policy, avoiding additional carbon emissions by not installing new modules unnecessarily, except when using second-hand panels with no embodied carbon. Conversely, in higher-budget scenarios, a more proactive installation policy results in higher carbon emissions over the system's lifetime, with the exception of the second-hand panel scenario. In this case, frequent updates increase the system's energy yield and consequently its carbon savings.

Mixed Reward

The inclusion of environmental balance in the agent's reward function clearly benefits the models, resulting in higher Net Carbon Savings (NCS) across all scenarios compared to the base policy (mean increase of 8.9%) and



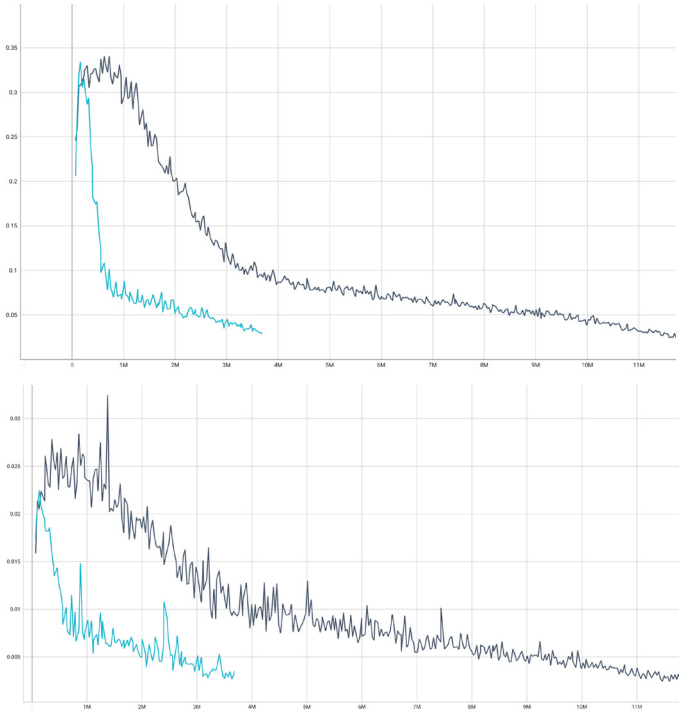


Figure 64: Training and evaluation metrics of complex (black) and simple (blue) action space based models. evaluation reward, policy gradient loss, clip fraction and approximate Kullback-Leibler divergence (measure of the difference between the new policy distribution and the old policy distribution)

the financial reward models (mean increase of 4.3% for low budget scenarios and 9.2% for high budget scenarios). However, this comes at a significant cost to Net Present Values (NPV). Given that financial payback periods are shorter than environmental payback periods (3-4 years versus 4-6 years), the inclusion of environmental objectives forces the agent to adopt a more conservative installation policy. This is especially evident when comparing the NPVs of these models against those based solely on financial rewards, resulting in a mean NPV decrease of 5.1%.

Most importantly, the viability of the multi-panel concept, where the agent chooses between new panels and second-hand modules (with no embodied carbon), is demonstrated. This approach yields the highest growth in carbon savings compared to both the base policy and financial balance-only models, while avoiding significant compromises in financial gains over the system's lifetime.

5.15.3 Model Performance

For a comparative analysis both models with complex and simplified action spaces for the low budget select panel scenario for house 2 were selected.

The model operating under a simple action space

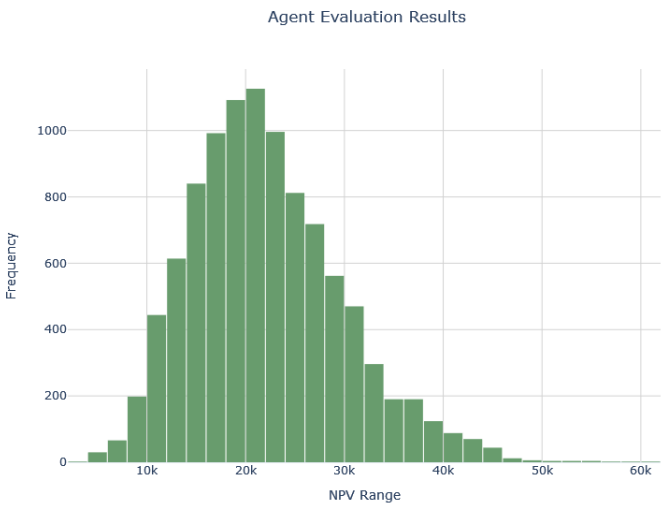


Figure 65: A) Bar Distribution of the NPVs for the model evaluation

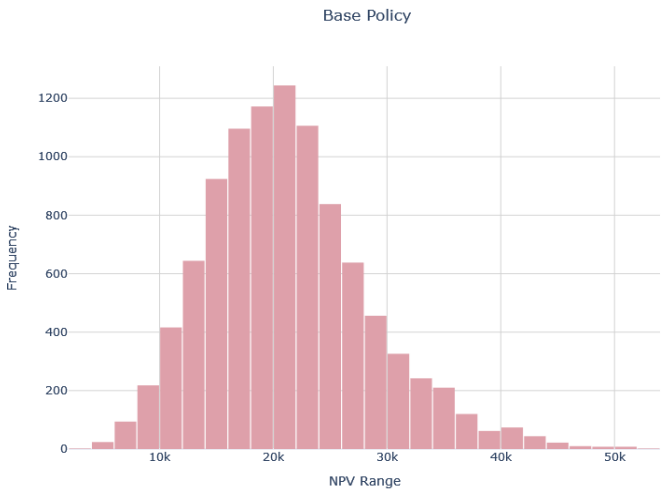


Figure 65: B) Bar distribution of the NPVs for the base-policy evaluation

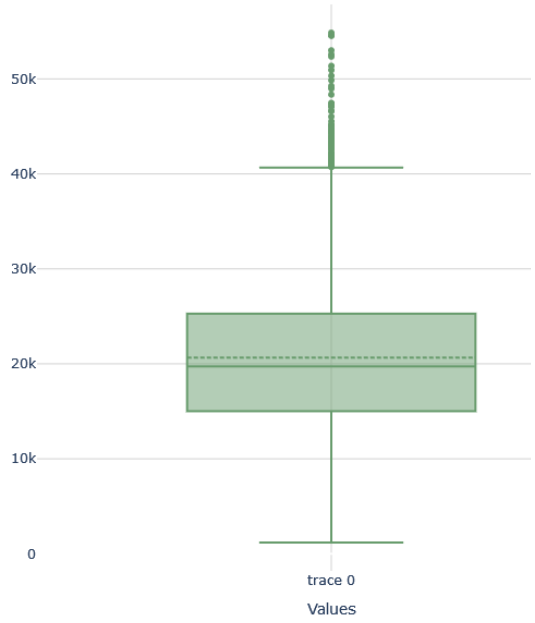


Figure 65: C) Box Plot of the NPVs for the model evalua-

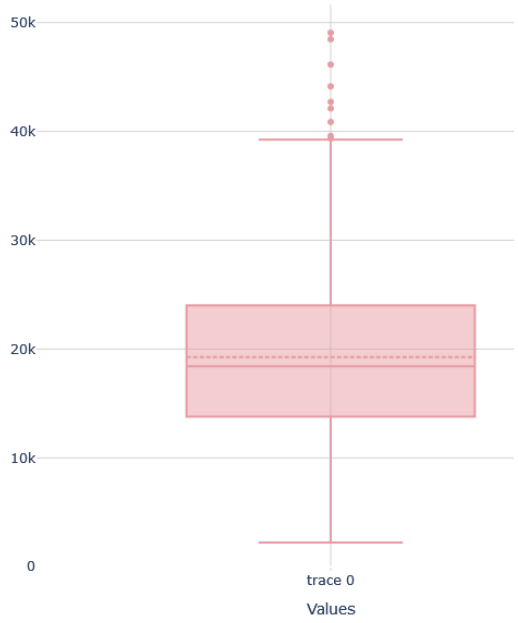


Figure 65: D) Box Plot of the NPVs for the base-policy evaluation

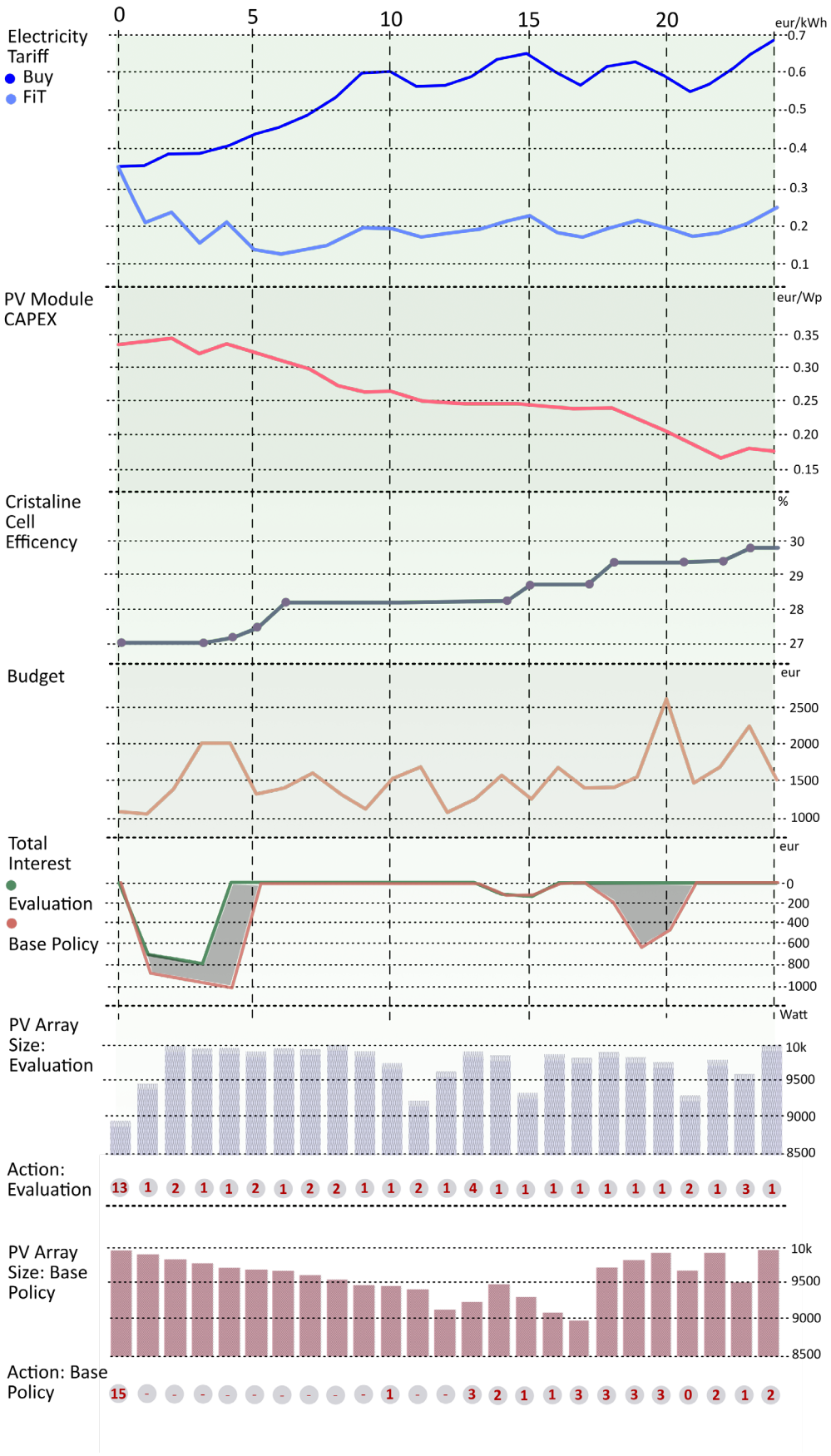
demonstrates accelerated and stable convergence across multiple parameters, including approximate KL divergence, clip fraction, policy gradient loss, and mean reward metrics. This rapid stabilization suggests that simpler action spaces facilitate quicker policy optimization and reduced variance in updates, likely due to a more constrained set of choices that allow faster generalization.

Conversely, the model with a complex action space shows a slower reduction in KL divergence and clip fraction, and a generally higher plateau in policy gradient loss. This indicates ongoing adjustments in policy updates and a longer trajectory towards optimization. The increased fluctuations observed in these metrics reflect the inherent challenges of navigating a larger decision space, where the optimization process must account for a wider variety of actions and potential outcomes.

5.16 Conclusion

In this chapter, a comprehensive optimization of the photovoltaic array was conducted. A workflow integrating all relevant variables identified during the literature review was developed and simulated using various stochastic and probabilistic methods. These variables were simulated using the tools described in Chapter 4, except for the PV module capital expenditure per Wp, where a Geometric Brownian Motion with a learning rate was employed.

05 PV OPTIMISATION



-	Achieved NPV
Model Evaluation	22 486
Base Policy	20 364

Figure 66: Comprahensive plot of a selected episode

05 PV OPTIMISATION

Following the introduction of each variable into the simulation environment, the algorithm's performance and output were evaluated. It was determined that modifying the action space compared to the simplified version used in the toy problem could provide benefits, and this revised action space was tested as an alternative.

The model's hyperparameters were then fine-tuned using manual adjustments and the Optuna library. The revised approach was tested across various scenarios involving different types of PV modules and household energy demands. The results were predominantly positive, with the trained model consistently outperforming the base policy. From the optimization, three primary conclusions were drawn:

1. The simplified action space yields better performance than the complex action space.

2. As roof size increases, allowing for more panel configurations, and as the budget decreases, the trained model increasingly outperforms the base policy.

3. The reinforcement learning (RL) framework effectively incorporates multiple objectives into the optimization process and accommodates more than one module type.

06. PV + BES System Optimisation

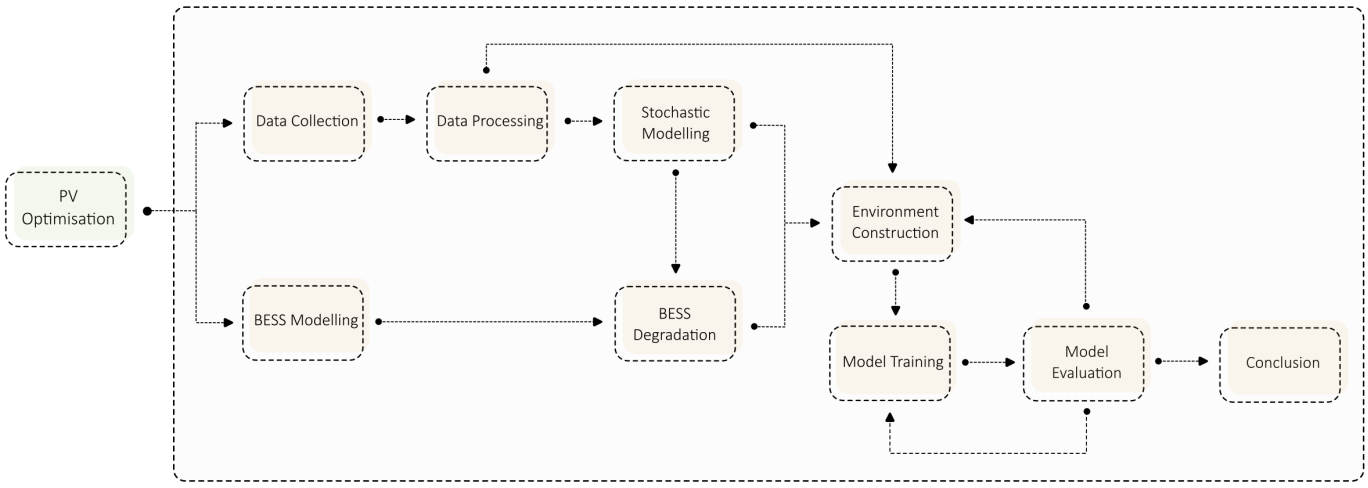


Figure 67: PV and BESS Optimisation - Flowchart

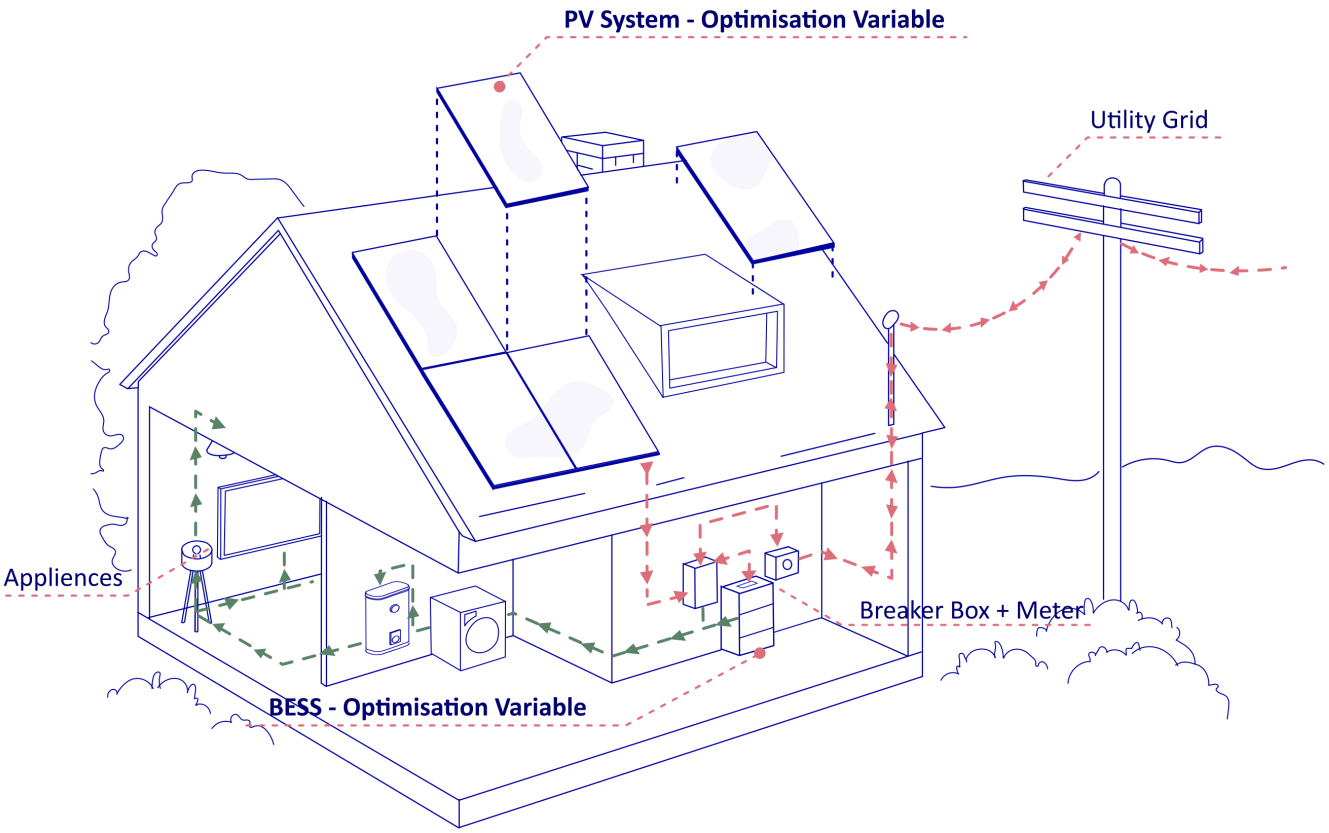


Figure 68: PV and BESS Optimisation

06 PV + BESS OPTIMISATION

This chapter focuses on the optimization of the battery energy storage system integrated with rooftop photovoltaic panels. The methodology is identical to that of the previous chapter concerning stochastic scenario generation, electricity tariffs, financial and environmental variables, household loads, and the reward function

6.1 Battery CAPEX

Similar to the methodology outlined in section 5.7 for photovoltaic price scenario generation, the approach for lithium-ion batteries involves employing geometric Brownian motion coupled with mean drift estimation. This method utilizes a learning rate derived from predicted cumulative installation capacity. As noted by Ziegler and Trancik (2021), simulating annual real prices scaled by energy capacity yields a high coefficient of determination. Nagelhout et al. have determined a learning rate of 17%, as assumed in this study, indicating that lithium-ion battery cell prices decrease by an average of 17% with a doubling of installed capacity, following Wright's Law.

Historical global prices of lithium-ion battery modules are utilized to compute the volatility of the geometric Brownian motion. On the other hand, Vartiainen et al., 2019 have simulated various scenarios for the market volume of BESS in Europe until 2050. Their findings are leveraged to estimate future lithium-ion battery storage prices. Following the approach outlined in section 5.7, the base scenario for the cumulative capacity value is selected.

Variable	Sigma	Mu	Theta
Battery CAPEX	0.122	-0.038	-

Table 24: Parameters of the Battery CAPEX GBM

6.2 Battery Modelling

Similarly to the PV system, it would be best to use a precise physics based cyclic model as the one described in Deng et al., 2022. However due to the time constraints it was decided to utilise a battery model presented in (Qi , Rashedi and Ardakanian, 2019), that provides a practical and accurate linear approximation of the physical properties of a lithium-ion battery. According to this model, the energy content of a lithium-ion battery changes over time in the following way:

$$E_{Batery}(h+1) = E_{Batery}(h)\left(1-\eta_{p,leak}\right) + \frac{\Delta E_{Batery}(h)}{\eta_{inv,B}} - T_u \times \eta_{c,leak} \quad \text{eq. 81}$$

$$\Delta E_{Batery}(h) = \begin{cases} T_u Q_{charge}(h)\eta_c & if\ charging \\ \dfrac{-T_u Q_{discharge}(h)}{\eta_d} & if\ discharging \end{cases} \quad \text{eq. 82}$$

where:
h is an hourly timestep,
$\eta_{inv,B}$ is the efficency of the AC to DC battery inverter
$E(h)$ is Energy content of the battery at times t in watt-hours (Wh),
$\eta_{p,leak}$ is the leakage rate per time unit as a fraction of the SOC,
$\eta_{c,leak}$ Constant leakage rate of energy, measured in Wh per unit time,
T_u Duration of each time slot (one hour),
η_c is the charging efficency,
η_d is the discharging efficency,
$Q_{charge}(h)$ charge power (W),
$Q_{discharge}(h)$ discharge power (W).

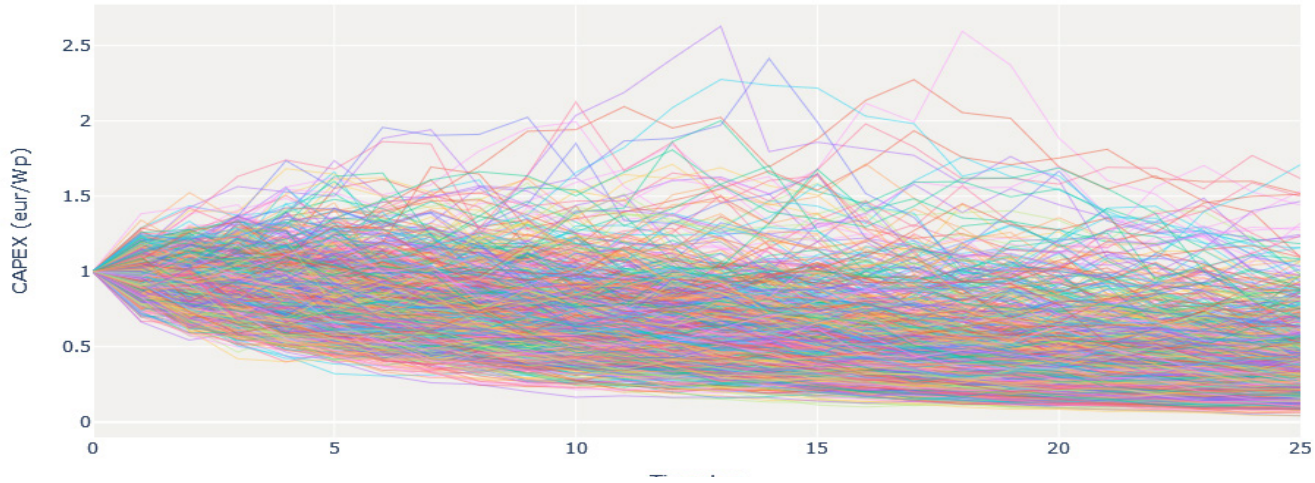


Figure 69: BESS CAPEX Scenarios

6.3 Battery Details

Lithium iron phosphate (LiFePO4) batteries currently dominate the residential home energy storage market, with no viable alternatives available. Furthermore, market research in the Netherlands indicates that purchasing used batteries is not a feasible option at present. Mainstream batteries in the market do not exhibit significant differences across various price ranges in terms of cycle durability, self-discharge rate, or depth of discharge. A mainstream mid-range Huawei Luna 2000, available in three different capacities ranging from 5 to 15 kWh, was selected for analysis. It allows for expanding the existing storage by 5kWh, with each module operating independently.

The analysed system is always AC-coupled, necessitating an additional inverter to mediate between the battery and the rest of the system. The inverter is assumed to have a conversion efficiency of 98%. System's architecture is presented on figure 71.

The parameters of the battery storage are set according to its specifications. Hence $Q_{charge,max}$ and $Q_{discharge,max}$ are set 1.5 kW, 3 kW and 4.5 kW for 5kWh, 10 kWh and 15 kWh modules respectively. The roundtrip efficiency is around 90% therefore charging and discharging efficiencies are set to 0.95. The recommended depth of discharge is 95% therefore SOC_{min} (minimal state of charge) is 0.05 x B_{cap} (battery capacity). On the other hand the recommendation for the maximal SOC is 100% for LiFePO4 batteries hence SOC_{max} = B_{cap}. The other parameters of the battery are set as follows: $\eta_{p,leak}$= 0, $\eta_{c,leak}$= $B_{cap} \times 10^{-4}$, T_u= 1.

6.4 Battery Degradation

Battery degradation and aging is crucial to consider in the system's optimisation. For consumer home BESS cycling aging is the dominant aging mode. The current condition of a battery is quantified by the State of Health (SoH), which measures the degree of capacity fade. A battery is considered to have reached its end-of-life when the SoH declines to 80% (Zhang et al., 2017). Operating the battery beyond this point introduces potential hazards due to increased risks of failure and safety concerns.

Typical LiFePO4 BESS can last from 10 to 14 years, under normal operating conditions, which accounts to about 4000 - 5000 cycles (Beltran, Ayuso and Pérez, 2020).

$$SoH = \frac{B_{cap,act} - B_{cap,EOL}}{B_{cap,nom} - B_{cap,EOL}} \times 100\% \qquad \textit{eq. 83}$$

where:

$B_{cap,act}$ is the current battery capacity
$B_{cap,nom}$ is the nominal capacity

$$B_{cap,EOL} = 0.8 \times B_{cap,nom} \qquad \textit{eq. 84}$$

The degradation curve of lithium-ion battery storage typically exhibits a non-linear pattern, where the capacity of the battery declines over time due to chemical and physical changes within the cells. This study uses a non-homogenous gamma process (NHGP) to simulate the degradation paths exhibiting monotonic, nonlinear, and non-reversible behaviors. This approach has been developed by Galatro et al., (2021) to account for cell-to-cell variations. NHGPs PDF is:

$$f_{a,b}(x) = \frac{1}{\Gamma(a\varphi(t))} b^{a\varphi(t)} x^{a\varphi(t)-1} e^{-bx} \qquad \textit{eq. 85}$$

where $\Gamma(a\varphi(t))$ is the Gamma function for the shape parameter; $\varphi(t) = c$ where c is a nonlinear constant and b is the scale parameter.

Ideally this process should be fitted to real data using its likelihood function. However due to times constraints the parameters were estimated empricially from several studies on LiFePO4 battery degradation (Johnen et al., 2021), (Sun et al., 2018), (Thomas et al., 2008).

According to Carpinelli et al. (2014) under time-of-use electricity tariff with three price levels, as considered in this study, a home BESS is typically subjected to 1 cycle a day. To model battery degradation more accurately, a rainflow algorithm has been incorporated into the analysis framework to compute the annual number of charging and discharging cycles of the battery. The rainflow algorythm processes a time series state of charge data of the battery and identifies stress-reversal points within the series. It quantifies the number of cycles, categorizing each based on the amplitude and mean of the cycle (Angenendt et al., 2018). A specialised *rainflow* library was used.

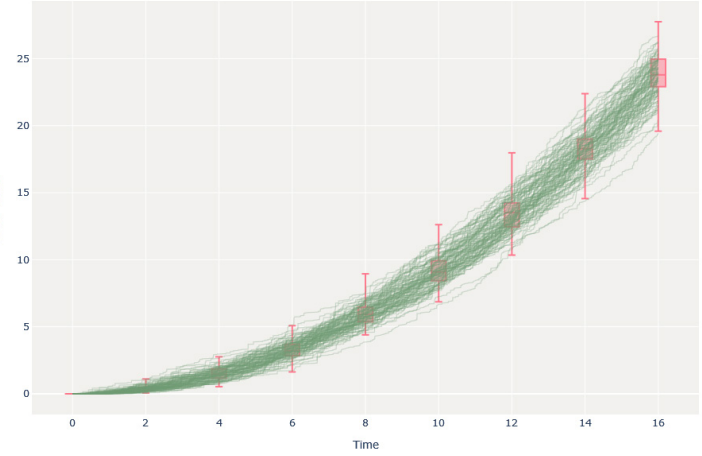


Figure 70: A plot of 100 NHGP for the battery degradation paths

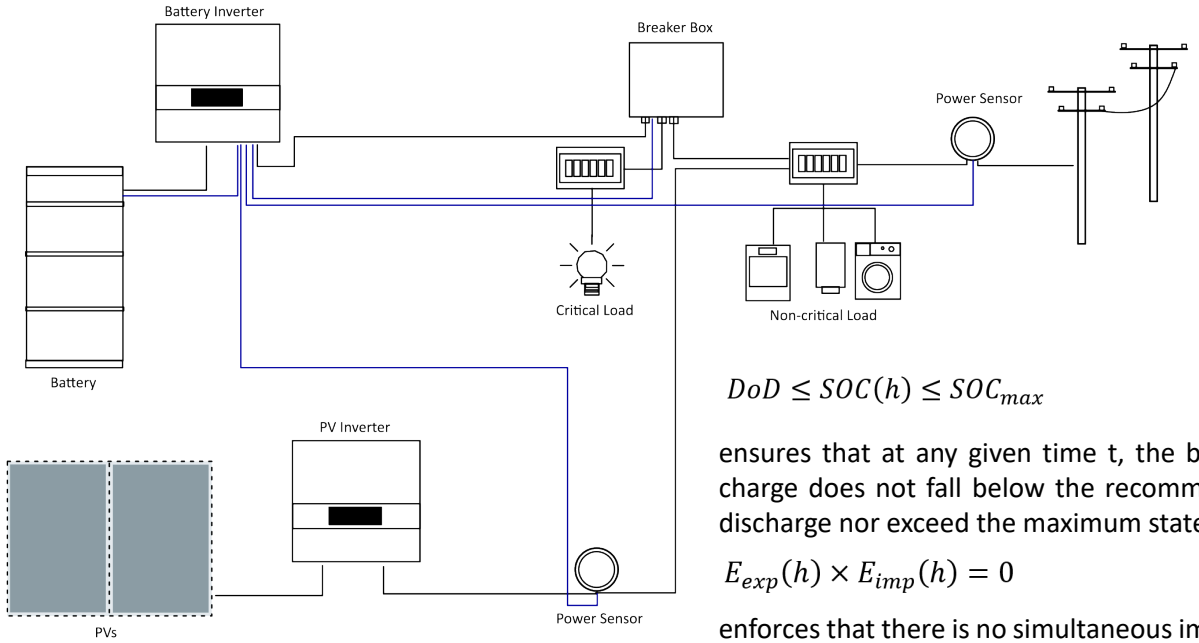


Figure 71: System under consideration

6.5 System Balance and Constraints

The system is required to conform to a series of constraints that ensure all components maintain equilibrium:

$$SOC(h) = \frac{E_{Batery}(h)}{B_{cap}} \quad \textit{eq. 86}$$

$$P_{load}(h) = P_{PV,total}(h) + E_{sup}(h) + E_{Batery}(h) \quad \textit{eq. 87}$$

The load is matched by total power sourced from the grid, generated by the solar system, and drawn from the battery.

$$P_{PV,total}(h) = P_{load}(h) + E_{exp}(h) + E_{Batery}(h) \quad \textit{eq. 88}$$

The electricity produced by the PV system is accounted for, either being used locally, stored for future use, or exported to the grid.

$$E_{exp}(h) \leq E_{exp,max} \quad \textit{eq. 89}$$

Ensures that the hourly energy exported to the grid is no higher than the maximum of $E_{exp,max}$ = 4kW.

$$Q_{charge}(h) \times Q_{discharge}(h) = 0 \quad \textit{eq. 90}$$

The battery is either charging or discharging, but not doing both simultaneously.

$$0 \leq Q_{charge}(h) \leq Q_{charge,max} \quad \textit{eq. 91}$$

$$0 \leq Q_{discharge}(h) \leq Q_{discharge,max} \quad \textit{eq. 92}$$

Ensure that the battery's charging and discharging adhere to the maximum rates allowable for the battery.

$$DoD \leq SOC(h) \leq SOC_{max} \quad \textit{eq. 93}$$

ensures that at any given time t, the battery's state of charge does not fall below the recommended depth of discharge nor exceed the maximum state of charge.

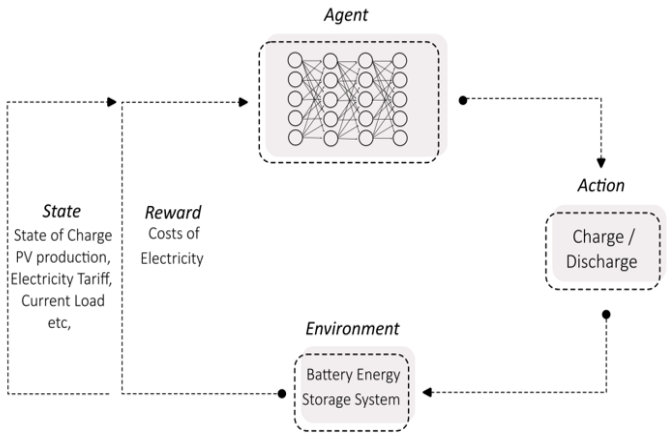
$$E_{exp}(h) \times E_{imp}(h) = 0 \quad \textit{eq. 94}$$

enforces that there is no simultaneous import and export of power to the grid.

6.6 System Operation (RL within RL)

In order to realistically assess the advantages of adopting BESS, various operational strategies outlined in section 2.3.10 of this thesis were analysed. It is well-documented that predictive control algorithms surpass rule-based methods by adeptly adapting to variations in PV and battery sizes, energy loads, and electricity tariffs. However, no metaheuristic algorithms were immediately available or applicable. Given the author's expertise in reinforcement learning, it was decided to develop a bespoke algorithm from scratch. Machine learning-based techniques have been demonstrated to achieve superior results compared to traditional methods (Corte Real et al., 2024). The methodology employed mirrors that of the planning model, with the environment constructed in the OpenAI Gym and an agent trained using Stable Baselines. As noted in the literature on model predictive control, it is typical to precede this with the development and training of a regression model responsible for predicting building load and weather conditions a few timesteps into the future. However, in our scenario, this step is unnecessary since the future values of all necessary parameters are known within the simulation. Hence, the methodology would need to be expanded for application to real-life scenarios.

In configuring a Battery Energy Storage System (BESS) operation algorithm, it's common practice to prioritize system longevity. While some authors incorporate battery aging models to minimize cycle count and extend lifespan, this project doesn't explicitly adopt such models. Instead, it focuses on constraints such as maximum charge



depth and state of charge to prevent excessive use and degradation. This approach, though simpler, has been shown to be equally effective in studies like Angenendt et al., 2018, particularly in evaluating the total levelized cost of electricity for the system.

The problem is formulated as an MDP optimal control problem over a finite time horizon. Initially, an episode length of 8760 time steps, corresponding to the entire year, was tested. However, optimal convergence couldn't be achieved due to the extended episode duration. Consequently, the episode length was reduced to 168 hourly time steps, covering a single week. During each timestep, the agent receives input regarding weekly tariffs, photovoltaic production, and building loads, which are randomly provided.

6.5.1 Battery System Environment

The observation space encompasses the state of charge at t and t-1, PV and house load balance at h, h+1 and h+2, the electricity tariff at h, Feed in tariff at h, hour of day (1 to 24), the week of year (1 to 52), B_{cap} and P_{max} of the PV panels:

$$OBS(h) = [SOC(h),\ SOC(h-1),\ N(h),\ N(h+1),\ N(h+2),\ I(h),\ J(h),\ Hour,\ Week,\ B_{cap},\ P_{max}] \quad \text{eq. 95}$$

where:

$$N(h) = P_{PV,total} - P_{load} \quad \text{eq. 96}$$

is the net hourly load.

All values in the observation space are normalised from 0 to 1. B_{cap} and P_{max} which are subject to variation, are individually adjusted for each case house. For house 1 B_{cap} ranges from 4 kWh to 5 kWh, P_{max} ranges from 2 490 Wp to 4 980 Wp.

Actions correspond to the power level of charging or discharging the battery, which is a interval that ranges from maximum discharging power ($Q_{discharge,max}$) to the maximum charging power of the battery ($Q_{charge,max}$) multiplied by the charging efficency.

$$A = [Q_{discharge,max}\ \ Q_{charge,max}] \quad \text{eq. 97}$$

This is normalised to from -1 to 1 in the actual environment, with discharging taking the negative values.

$$A = [-1, 1] \quad \text{eq. 98}$$

Upon receiving the action A from the agent, the environment checks whether the current time step t is not the final one, t≠H. If this condition holds, it proceeds with the following updates (figure 72).

Model's reward is the hourly financial balance of the expense for the imported energy and revenue for the exported energy.

$$R = [-I(h) \times E_{sup}(h) + J(h) \times E_{exp}(h)] \quad \text{eq. 99}$$

6.5.2 Training and Hyperprarameters

Similarly to the main problem here PPO is also used based on the very recent findings of Corte Real et al., 2024. An alternative policy using an LSTM (Long Short-Term Memory) network, rather than a simple neural network, was explored. LSTMs are a type of recurrent neural network (RNN) well-suited for processing sequences and time-series data, making them advantageous for dynamic environments where past information is crucial for current decisions. In our case the use of an LSTM could be beneficial as it enables the model to retain information over extended periods, enhancing the agent's ability to predict and optimize energy usage based on historical usage patterns and variable load demands. However, initial tests not only did not indicate that LSTM-based policy improves the model's convergence, but furthermore the training was slowed down significantly. Therefore, a standard NN was used.

To tune the hyperparameters of this model a similar methodology to the one emlpyed to the main model, in the last chapter was applied. However, in order not to bore the reader the description of this step is ommited. The agent is trained using multiprocessing on 16 parallel environments. Learning is set to 0.0003. The discount factor, gamma, was set to 0.99. The lambda parameter in was set at 0.95. The numbers of epochs,

```
# CONDITION A - load is HIGHER than pv production (there is a shortge)
if self.step_pv_load < 0:
    if action < 0: # CONDITION A1 - the batery Discharges

        self.step_pv_load_plus_batery_discharge = self.step_pv_load + actual_discharge
        self.current_SOC -= actual_discharge

        if self.step_pv_load_plus_batery_discharge > 0: # CONDITION A1A - the energy shortage is satisfied by batery,
            to_grid += min(self.step_pv_load_plus_batery_discharge, self.max_export) # then any surplus to grid.

        else: # CONDITION A1B - the energy shortage is NOT satisfied by batery,
            from_grid += (- self.step_pv_load_plus_batery_discharge) # then shortage supplemented from grid

    elif action > 0: # CONDITION A2 - despite the shortage the batery charges

        self.current_SOC += actual_charge
        from_grid += (- self.step_pv_load) # then shortage is imported from the grid,
        from_grid += actual_charge # and energy is imported for charging.

else: # CONDITION B the load is LOWER than pv production (there is a surplus)

    if action >= 0: # CONDITION B1 - the batery Charges
        self.step_pv_load_minus_batery_charge = self.step_pv_load - actual_charge
        self.current_SOC += actual_charge # charge the batery

        if self.step_pv_load_minus_batery_charge > 0: # CONDITION B1A - the surplus persists after charging batery,
            to_grid += min(self.step_pv_load_minus_batery_charge, self.max_export) # then it is sold to the grid

        else:   # CONDITION B1B - the charge is higher than surplus
            from_grid += ( - self.step_pv_load_minus_batery_charge)

    if action < 0: # CONDITION B2 - Despite there being a surplus the batery discharges
        self.current_SOC -= actual_discharge
        to_grid += self.step_pv_load + actual_discharge
```

Figure 72: Transitions of the BESS operational environment

batches, and timesteps were set at 16, 2048 (128 per single environment), and 15 000 000 respectively. PPO's network use NN architecture consisting of five fully-connected layers with 2048, 1024, 1024, 512, and 256 neurons, respectively, using the tanh activation function. This is presented on figure 73.

6.5.3 Evaluation

The results of the model are compared against a simple rule-based operation method from Ngoc An et al., 2015. This operation strategy involves the battery charging if the net hourly load $N(h) < 0$, meaning that there is a shortage within the system and discharging if $N(h) > 0$, which indicates an energy shortages. All relevant operation costraints from section 6.4 apply.

6.5.4 Results

The model was succesfully trained for house 1. Overall

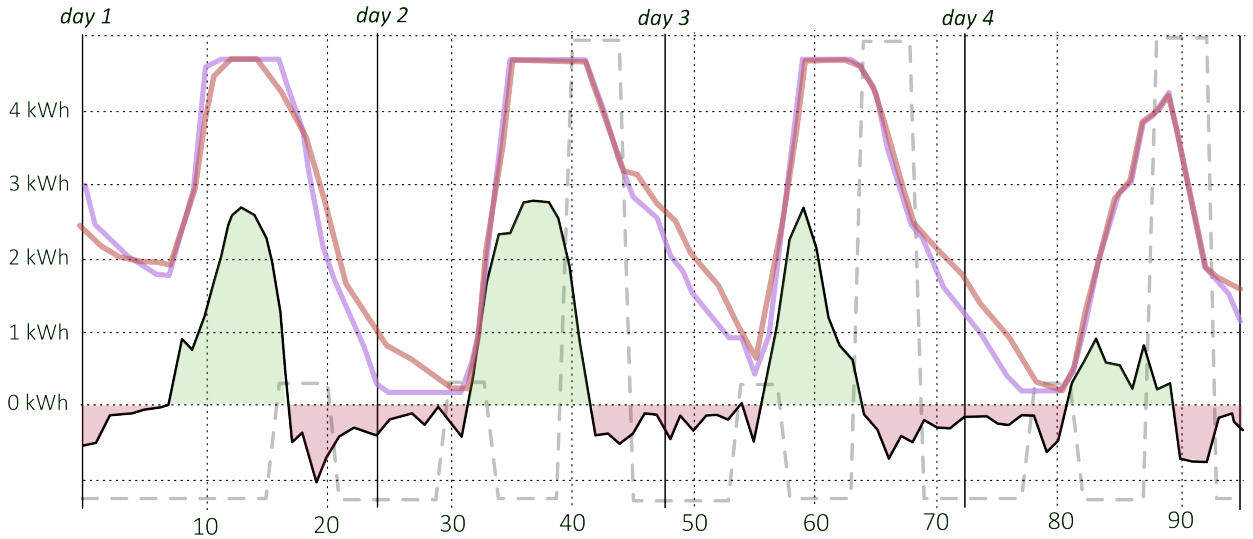


Figure 74: SoC of the Battery plotted for April 22nd to April 26th (B_{cap} = 5 kWh, $P_{PV,max}$ = 4 980 Wp), red line is the RL operated BESS, purple is the rule-based policy operation, dashed line indicades the electricity tariff rates, red and green fields indicate electricity shortage or surplus from PV and building load .

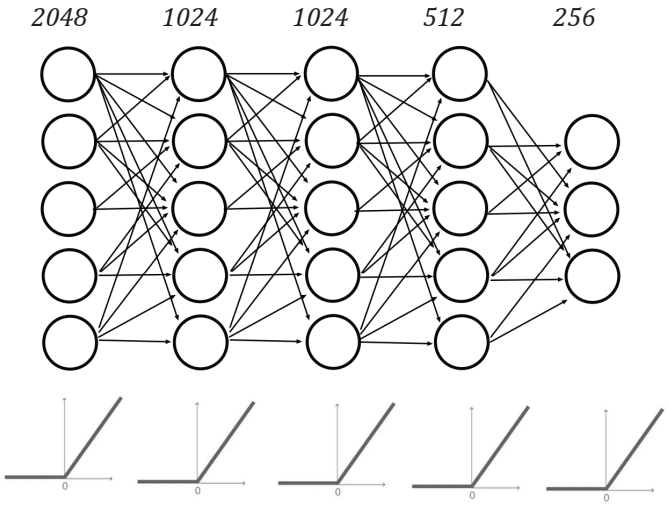


Figure 73: Network Architecture for the BESS operation PPO's policy

its performance surpases the performance of the rule-based method. Figure 67 presents the mean reward from 10,000 evaluation episodes, alongside results from a rule-based policy and a scenario without battery storage. The trained agent slightly outperforms the rule-based policy, achieving -108.9 EUR compared to -121.4 EUR. However, both strategies significantly surpass the performance in the absence of battery storage, which results in -949.7 EUR. The operation of the battery under these two methods is further elucidated in Figure 74, which illustrates the battery's SOC. It is evident that the battery managed by the model discharges more slowly, particularly during periods of low tariffs, thereby preventing the battery from reaching minimal SOC and allowing for a more evenly distributed discharge over extended periods. The training and evaluation metrics for this models are depicted in Appendix L.

6.5.5 Implementation

The implementation of the trained model for battery operation into the main environment resulted in a substantial slowdown in training, likely due to the size of the neural networks within the model. This led to a training slowdown of more than 25 times compared to similar models discussed in the previous chapter. This significant decrease in training speed was deemed unsuitable given the time constraints of this study. While it was considered that a regression model could be developed to approximate the outputs from the battery operation model and then integrated into the optimization environment, this approach was not pursued due to similar time constraints. Moreover, since the benefits of using the trained model over the rule-based method were not very significant, the rule-based method was implemented in the optimization environment. This resulted in a 6x slowdown in training relative to the environments evaluated in previous chapter.

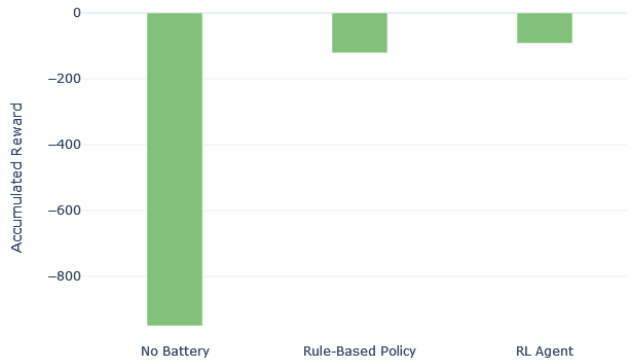


Figure 75: Comparison of the trained agent operation to the rule-base policy and scenario with no BESS.

6.7 Action Space

The action space remains largely the same as in the previous chapter. Similarly to that, two distinct action spaces are being utilized. The **simplified action space** is as follows:

$$A = \{(x, y, z) \mid 0 \leq \mathrm{x} \leq \mathrm{N} \text{ and } 0 \leq \mathrm{y} \leq \mathrm{M} \text{ and } 0 \leq \mathrm{z} \leq \mathrm{L}\} \quad \textit{eq. 100}$$

where:
x is the number of panels exchanged,
y is the type of panel chosen,
z is the size of BESS added,
N is the maximum number of panels,
M represents the maximum number of panel types.
L is the maximum no of BESS units.

If z is one a 5kWh module is added, if 2 - 10kWh if 3 - 15kWh. Note that if the maximal size of the BESS is installed ($B_{cap,max}$), then action z has no influence on the environment etc.

$$B_{cap} \leq B_{cap,max} \quad \textit{eq. 101}$$

In the preliminary calculations, $B_{cap,max}$ for each house was determined. For instance, it was found that for House 1, the benefit of installing a BESS larger than 5 kWh is marginal; thus, larger options were excluded from the optimization. The $B_{cap,max}$ for each house is displayed in Table 25.

Similarly, a single integer has been added to the **complex action space**.

$$A = \{(x_i, x_{i+1} \ldots, x_n, y, z) | 0 \leq x_i \leq 1 \; and \; 0 \leq y \leq M \; and \; 0 \leq z \leq L\} \quad \textit{eq. 102}$$

The observation space has also been minimally adjusted to include measures for the current price of battery, and itscapacity multiplied by its state of health (SoH).

$$OBS(t) = [\eta_i(t), \eta_{i+1}(t), \dots \eta_{i+n}(t), I(t), J(t), Eff(t),\ CAPEX_{PV}(t),$$
$$C(t-1), Pen(t), B(t), B(t-1), EG(t), id_i(t), id_{i+1}(t), \dots id_{i+n}(t),$$
$$B_{cap}(t) \times SoH(t),\ CAPEX_{BAT}(t)] \qquad \text{eq. 103}$$

where:

$$C_{BAT}(t) = C_{module}(t) + Inst_{BAT}(t) + BoS_{BAT}(t) \qquad \text{eq. 104}$$

Values for C_{module} and $Inst_{BAT}$ at *t=1* are given in table 25.

$Inst_{BAT}$ is adjusted for inflation. C_{module} is multiplied by the current $CAPEX_{BAT}$. Disposal costs have not been included in the analysis, as it is uncertain whether such costs will be incurred. Additionally, with anticipated advancements in recycling technologies, disposal activities could potentially yield benefits rather than impose costs.

6.8 Observation Space

Contrary to other equations and components of the environment, the reward structure remains identical, as explained in section 5.12. Due to unavailabilty of data on environmental impact of residential BESS in the european region, in this chapter we only focus on the financial reward.

The training parameters, and agent's hyperparameters are the same compared to the last chapter.

	House 1	House 2	House 3
Max Battery Size	*5kwh*	*5kwh*	*10 kwh*

Table 25: Maximal BESS sizes for each House

C_{BAT} - 5kWh	3000 eur	-
C_{BAT} - 10kWh	5000 eur	-
C_{BAT} - 15kWh	7000 eur	-
$Inst_{BAT}$	1000 eur	-
BOS_{BAT}	50% of C_{BAT}	Only for initial installation

Table 26: Battery Installation Costs at t=1

```
#  ***
action_battery = action[1]
action = action[0]

# Find indices of the lowest 'action' values in previous_observation
indices = np.argsort(self.previous_observation[:self.number_of_panels])[:action]

# Replace these indices in the observation with efficiency_develop
self.observation[:self.number_of_panels][indices] = self.efficency_develop

# Copy over the other values from previous_observation to observation
mask = np.ones(len(self.previous_observation[:self.number_of_panels]), dtype=bool)
mask[indices] = False
self.observation[:self.number_of_panels][mask] = self.previous_observation[:self.number_of_panels][mask]

replaced_panels = np.zeros(len(self.previous_observation[:self.number_of_panels]), dtype=int)
replaced_panels[indices] = 1

instaltion = (self.observation[:self.number_of_panels] > 0).astype(int)
self.pv_costs -= instaltion.sum() * current_operational_costs

actions_step = np.array(replaced_panels)

if action_battery == 1:
    self.observation[31] = 1.
    self.battery_state = 1
```

Figure 76: Observation update in the RL enviornmnent, under the simplified action space

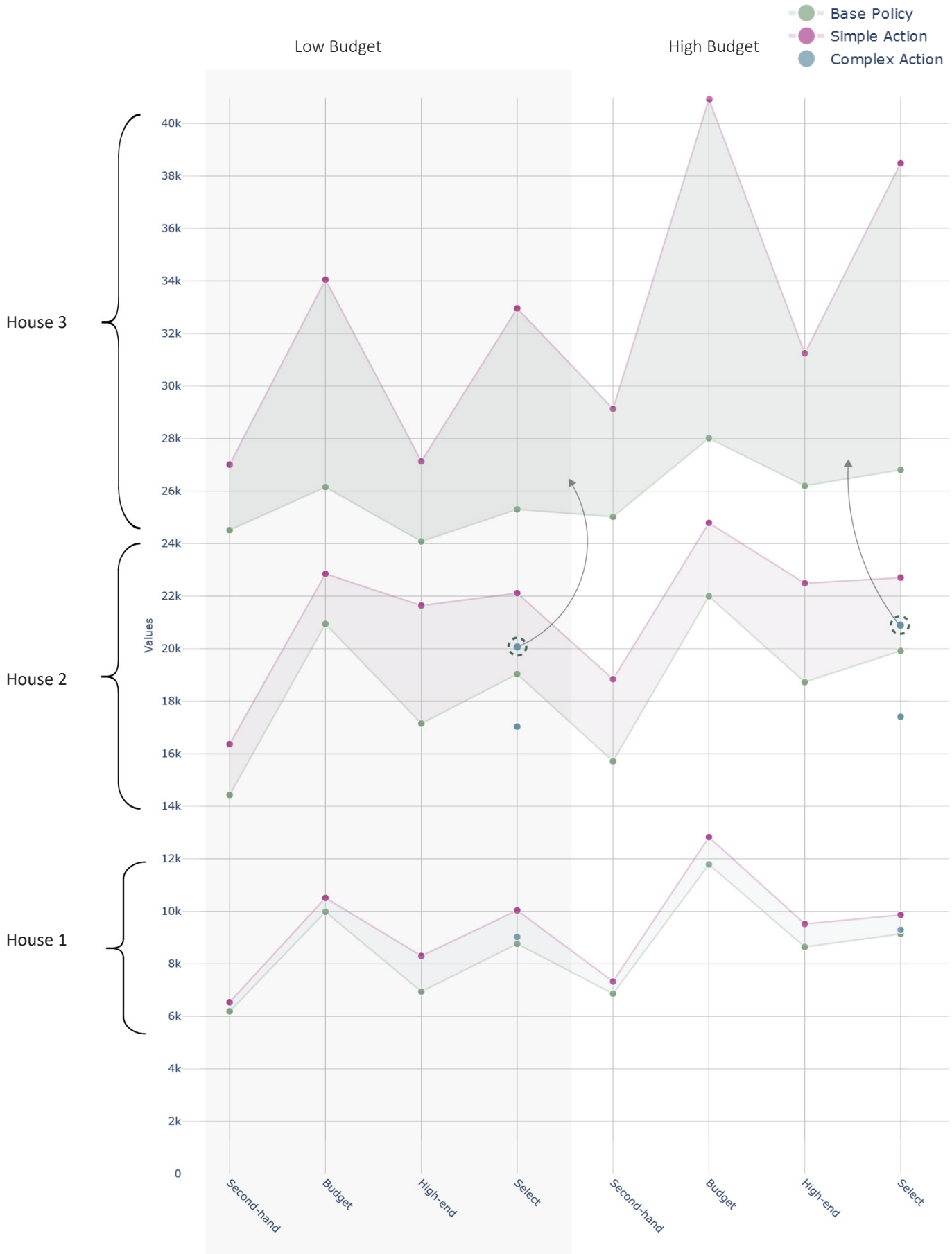


Figure 77: Comprahensive Plot of the NPV results for each analysed scenario under financial balance reward only

		Low Budget			High Budget		
		Base Policy	*Action Space*		*Base Policy*	*Action Space*	
Action Space		-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	6 192	-	6 538 (+5.6%)	6 865	-	7 325 (+6.7%)
Budget Panel	*NPV*	9 980	-	10 508 (+5.3%)	11 787	-	12 824 (+8.8%)
Hi-End Panel	*NPV*	6 946	-	8 300 (+19.5%)	8 645	-	9 518 (+10.1)
Select Panel	*NPV*	8 762	9 024 (+2.0%)	10 032 (+14.5)	9 139	9 294 (+0.7%)	9 861 (7.9%)

Table 27: A) Evaluation Results for House 1 - NPVs

		Low Budget			High Budget		
		Base Policy	*Action Space*		*Base Policy*	*Action Space*	
Action Space		-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	14 430	-	16 363 (+11.3%)	15 712	-	18 838 (+19.9%)
Budget Panel	*NPV*	20 944	-	22 849 (+9.1%)	21 998	-	24 791 (+12.7%)
Hi-End Panel	*NPV*	17 150	-	21 643 (+26.2%)	18 727	-	22 491 (+20.1%)
Select Panel	*NPV*	19 033	17 034 (-11.5%)	22 116 (+16.2%)	19 919	17 409 (-14.6%)	22 707 (+14.0%)

Table 28: A) Evaluation Results for House 2 - NPVs

		Low Budget			High Budget		
		Base Policy	*Action Space*		*Base Policy*	*Action Space*	
Action Space		-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	*NPV*	24 517	-	27 013 (+10.2%)	25 022	-	29 132 (+ 16.4%)
Budget Panel	*NPV*	29 154	-	34 052 (+16.8%)	31 797	-	39 926 (+25.5%)
Hi-End Panel	*NPV*	24 087	-	27 135 (+12.6%)	26 202	-	31 047 (+18.5%)
Select Panel	*NPV*	25 308	19 077 (-32.6%)	32 955 (+30.2%)	26 811	20 880 (-28.4%)	35 486 (+24.5%)

Table 29: A) Evaluation Results for House 3 - NPVs

6.9 Results

The outcomes of the model evaluation, specifically the Net Present Values, are detailed in Tables 27, 28 and 29, where they are compared against the base policy. The plot of NPVs for each analysed scenario is shown on figure 77. The NPV distributions for the scenario involving a budget panel with a high budget for house 2 are depicted in Figure 78. Additionally, a detailed plot for a selected episode within this scenario is illustrated in Figure 79. The results for each scenario are discussed below.

6.7.1 House 1

For the 2nd hand module scenario, in contrast to its equivalent with only photovoltaic systems discussed in Chapter 5, the trained model in this study outperforms the base policy, which dictates the installation of the BESS only and always when it is not installed. This leads to significantly higher interest costs at the beginning of the episode across all scenarios. Specifically, the mean interest cost for the low budget scenario exceeds 280 EUR for the second-hand panel, 370 EUR for the budget panel, and 600 EUR for the high-end panel, compared to zero interest costs for the second-hand panel during model evaluation (Appendix M). In all scenarios, the installation of the battery is postponed and not performed in the early steps. Under a low budget, the median timestep for BESS installation is 8 for the used panel, 12 for the budget panel, and 16 for the high-end panel. Additionally, in 26% of evaluated episodes for the high-end panel, the agent does not install the BESS at all. This percentage decreases to 12% for the budget panel and less than 1% for the second-hand panel. The model also tends to act more conservatively than the base policy by replacing the budget panel, under low budget, mainly at 77% efficiency. It is noteworthy that in fewer than 0.1% of the analyzed episodes across all scenarios, the battery was replaced while still functional.

When comparing complex to simplified action space-based models in the multi-panel scenario, the results are consistent with those from the previous chapter, with the simplified action space yielding superior results. This is the only case where the complex action-space-based model outperforms the base policy. However, this should be attributed to the inadequacy of the base policy rather than the commendable performance of the model.

6.7.2 House 2

As observed in Chapter 5, the performance difference between the base policy and the trained model increases. Similar to House 1, the highest benchmarking against the base policy is achieved in the high-end panel, low budget scenario. In this scenario, the model avoids the installation of all possible modules in 24% of evaluated scenarios and installs the battery in 68% of them. Typically, this installation occurs after the 18th timestep.

Figure 62 displays a selected episode for the budget panel in a high budget scenario. It is evident how the model manages to reduce additional interest costs compared to the base policy. In this case, the battery is only installed at timestep 15, when its capital expenditure reaches approximately 65% of its initial value.

The distributions of NPVs for the base policy and the model with simplified action space for the budget panel were analyzed using histograms and box plots. The histograms in Figures 61: A and B display a bell-shaped distribution for both the model evaluation and the base-policy NPVs, suggesting normal distributions. Figure 6.1: A shows a model evaluation range from about 20 000 to 28 000, with most NPVs clustered around 24 000, whereas Figure 6.1: B for the base-policy reveals a tighter distribution ranging from 12 000 to 25 000. The box plots in Figures 6.2: C and D further outline the spread and central tendency of NPVs. The presence of outliers in Figure 6.2: D suggests some variability under the base-policy evaluation.

The distributions of NPVs for the base policy and the model with simplified action space for the budget panel were analyzed reaveling differences in the central tendency and variability of NPVs between the two policies. Specifically, the model evaluation shows a higher median NPV and a tighter interquartile range as depicted by its box plot (Figure 6.1: C), indicating a more consistent and more favorable financial outcome compared to the base policy. Conversely, the base policy exhibits a wider range of NPVs with a lower median, as illustrated by the broader spread in both its histogram (Figure 6.1: B1) and box plot (Figure 6.1: D).

6.7.3 House 3

This presents the most complex observation and action spaces across all environments examined so far. The evaluation patterns further prove that the trained models outperform the base policy across all scenarios with the exception of the complex action space based models.

The cummulative mean total episode interests and the mean total electricity productions in kWh are included in Appendix M.

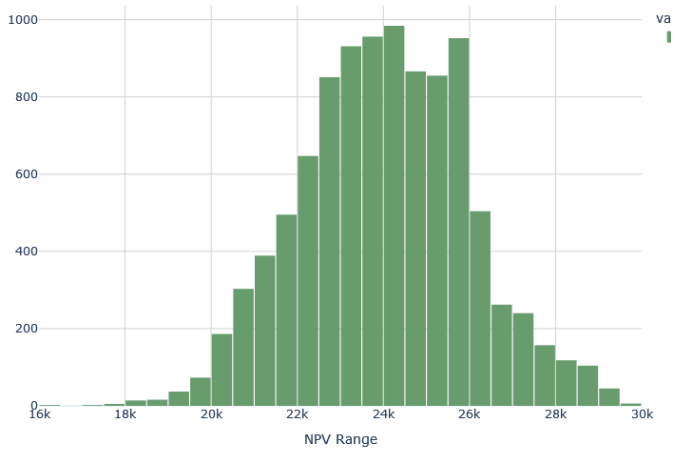


Figure 78: A) Bar Distribution of the NPVs for the model evaluation

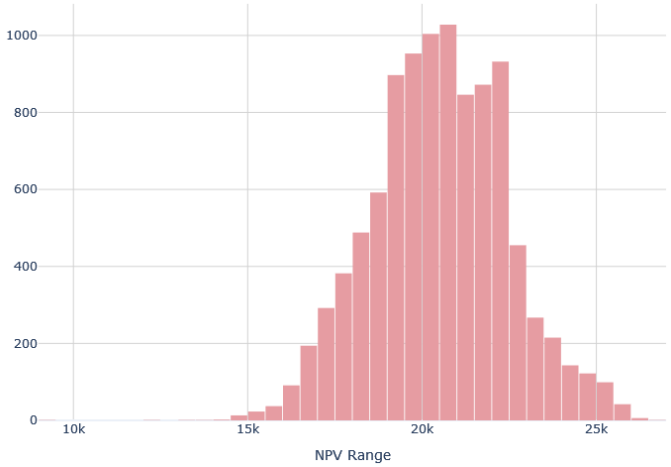


Figure 78: B) Bar Distribution of the NPVs for the base-policy

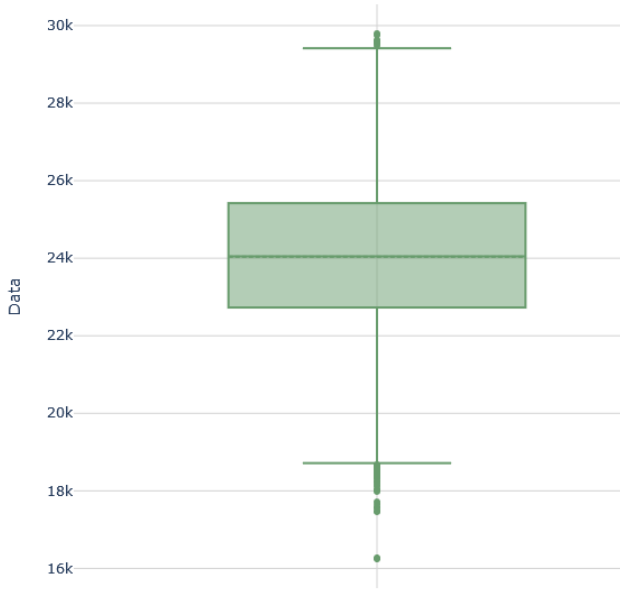


Figure 78: C) Box Plot of the NPVs for the model evaluation

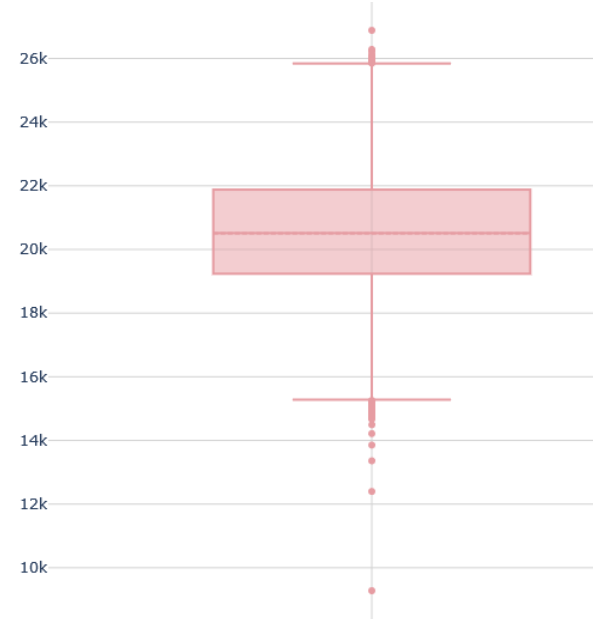


Figure 78: D) Box Plot of the NPVs for the base-policy evaluation

6.10 Conclusion

This chapter concentrates on optimizing a battery energy storage system integrated with rooftop photovoltaic panels. The methodology parallels that of the previous chapter, encompassing stochastic scenario generation, electricity tariffs, financial and environmental variables, household loads, and the reward function. It has been expanded to include critical variables related to the operation and installation of a BESS, featuring the development of a novel BESS operation algorithm based on Reinforcement Learning.

Similar to the findings in the previous chapter, the results here are predominantly positive, with the trained model consistently surpassing the base policy. This chapter reinforces two primary conclusions drawn in Chapter 5:

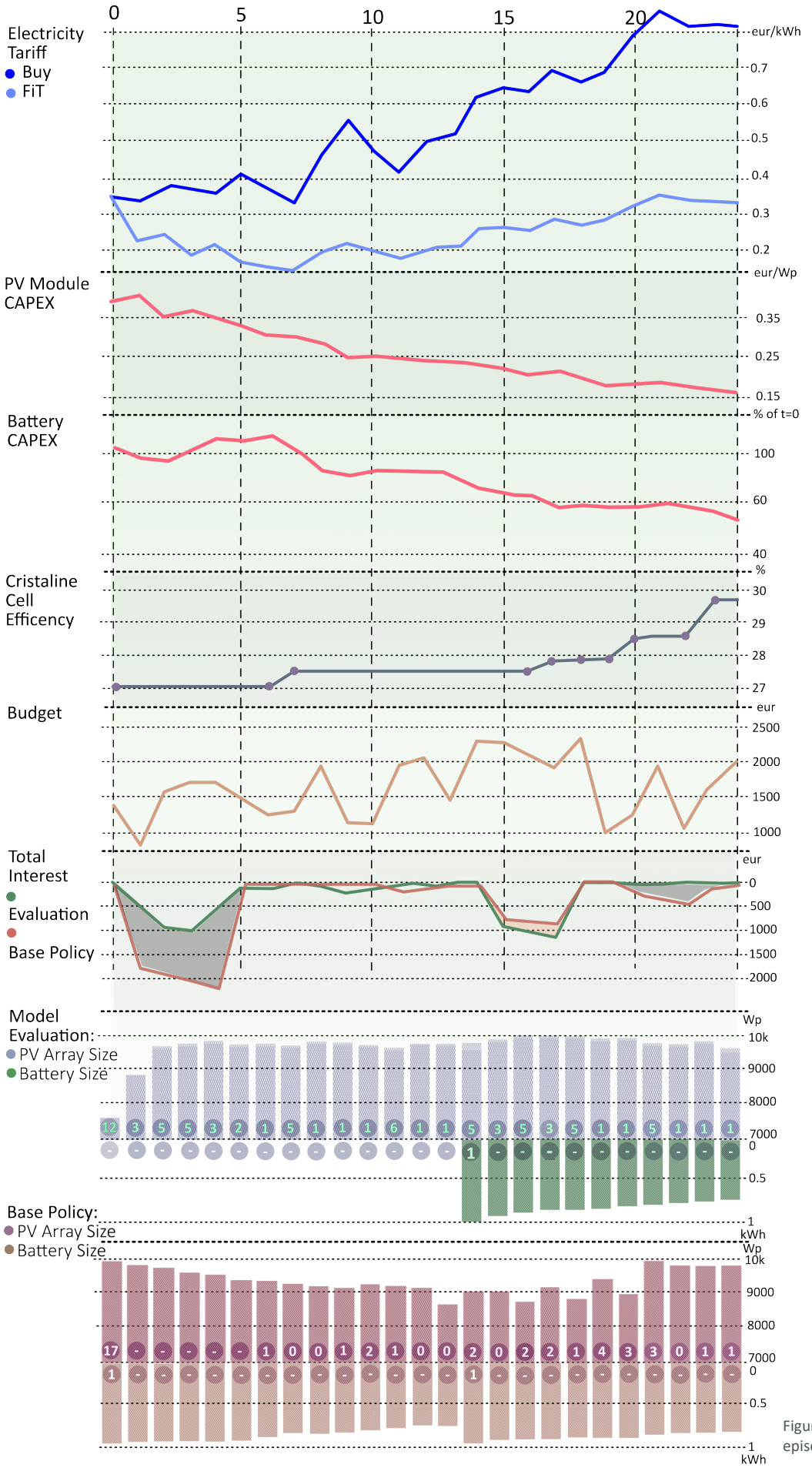
1. A simplified action space results in superior performance compared to a complex action space.

2. As roof size increases, allowing for a broader range of panel configurations, and as the budget is reduced, the trained model increasingly surpasses the performance of the base policy.

Furthermore, a third conclusion is articulated:

3. Integrating the BESS as a second optimization variable alongside the PV array enhances the model's performance relative to the base policy.

06 PV + BESS OPTIMISATION



-	Achieved NPV
Model Evaluation	30 331
Base Policy	26 342

Figure 79: Comprahensive plot of a selected episode

07. Deployment

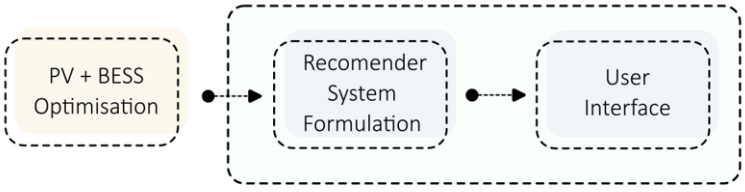


Figure 80: Deployment - Chapter Flowchart

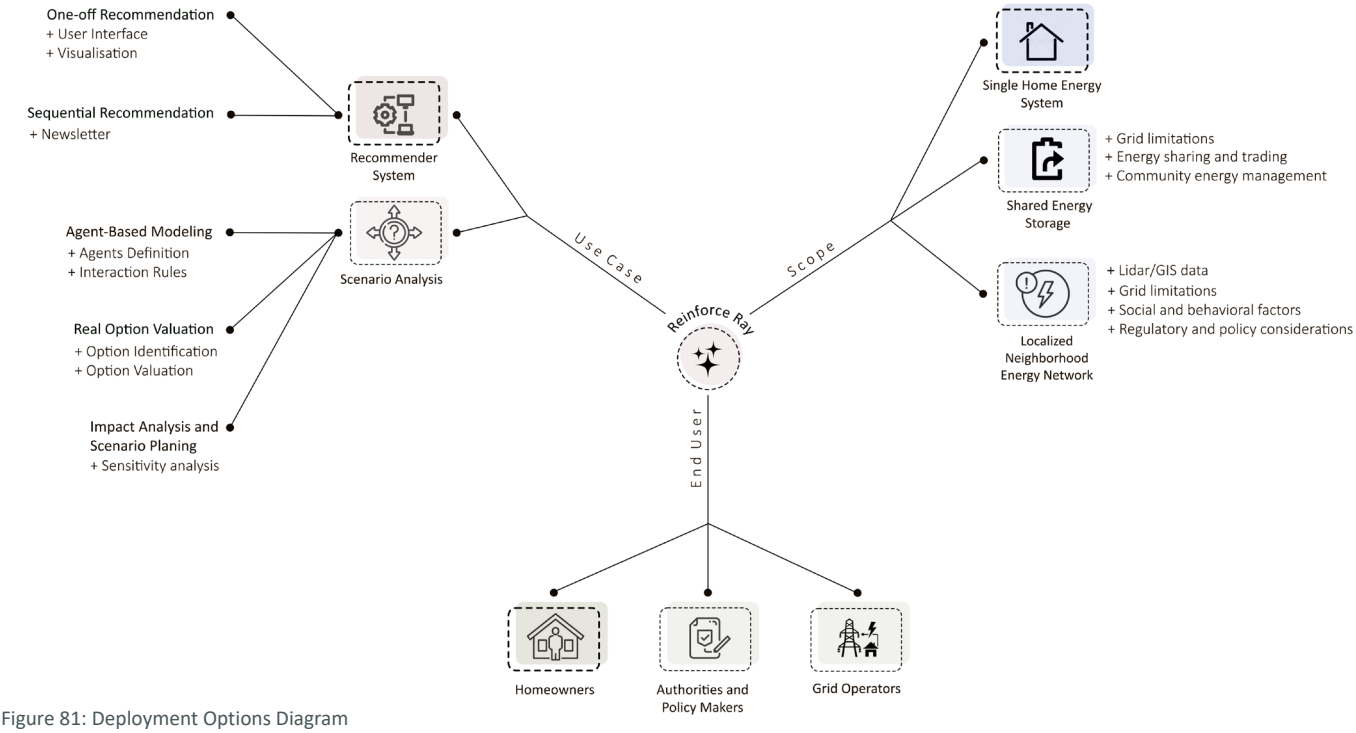


Figure 81: Deployment Options Diagram

7.1 Deployment Posibilities

Figure 81 presents an overview of potential deployment options for the RL models developed in this study. In general, the model's utility can be categorized into two main applications: it can either recommend the optimal action at any given time or provide foundation for scenario analysis.

Recommendations

This former option forms the basis for the recommender system development described in the subsequent section. The guidance provided by the RS can be either provided on an ad-hoc basis, requested periodically, or offered cyclically. It could also consist of a series of future recommendations based on both current and potential future states supplying the customers with flexible and responsive strategies. This can be effectively utilized by individual homeowners seeking to install or upgrade their home energy systems, small communities planning to deploy shared Distributed Energy Resources (DERs) like BESS, or companies leasing out their systems to individual users.

Moreover, when considering other factors such as electricity grid constraints (e.g., overvoltage, thermal limits) and spatial GIS data, the developed framework could be applied by municipalities and/or grid operators to assess the installation of larger-scale DERs in neighborhood or district-level low-voltage networks.

Scenario Analysis

An alternative implementation of the trained model could be to use it for a quantitative analysis to assess the value of different strategies under uncertainty. The model could aid in exploring and evaluating the evolution scenarios of influencing factors over time. This capability would be beneficial for residential customers, enabling them to manage their resources more effectively by leveraging their knowledge and best judgment on the future states of variables.

More realistically, this approach could significantly enhance its utility for local authorities and policy-makers in simulating various regulatory scenarios and market conditions to assess their influence on PV and BESS adoption rates and overall grid stability. Integrating RL models into Agent-Based Modeling (ABM) (Lee and Hong, 2019) frameworks, where each agent adheres to specific rules and interacts with others and the environment, allows for modeling adoption and usage patterns of these technologies among residential users. RL models empower agents to learn and adjust their strategies based on experience, environmental factors, and financial incentives. This also could be helpful for grid operators, allowing them to predict future energy demand patterns,

peak load times, and the potential need for grid upgrades or expansions.

The model can also be used to evaluate the general trends in the relevant factor's evolution and their impact on the optimal planning. More generalised scenarios for the electricity price or module CAPEX can be developed (ex. baseline, low, high) and then used for comparative analysis to identify trends, opportunities, and risks.

Furthermore, the framework can be useful for Real Option Valuation (ROV), which is a financial methodology that evaluates investment opportunities considering the flexibility to make future decisions based on how uncertainties unfold (Penizotto et. al, 2019).

It proves especially valuable for investments characterized by substantial uncertainty, such as DERs. These investments often involve partial or complete irreversibility, allowing for flexibility in project execution decisions and the potential to gather new information on critical variable developments over time. Wtihin ROV one can decide whether to install additional capacity, delay the decision, or alter the scale of installation (Andreolli, D'Alpaos and Moretto, 2021). This would further require to identify different options as well as valuating them against a base scenario. RL model can then calculate the payoffs of exercising different real options.

7.2 Reccomender System

The final phase of the proposed workflow involves developing a front-end recommendation system for the model. This system serves as a demonstrative interface, which could be integrated into an actual PV monitoring system. The main goal of the interface is to inform the end-user about the most - optimal range of actions to their PV system at a given time. The RS could be employed on both

Figure 82 gives an overview of reccomender system operation.

The software's API would automatically collect all necessary information for determining actions, including current energy prices, PV modules, and battery energy storage systems. The API would scan current offers for system components from retailers, installers, and manufacturers. Additionally, it would gather data regarding the current state, size, and installation locations of the existing devices from the monitoring systems or IoT devices for the PV array and battery storage to make real-time decisions .

Regarding the described fundamental disctinction between deployment methods given in 2.5.1. This RS can be both implemented offline - meaning that the recomendations would be given periodically to the user ex. as a "newsletter" sent them via email or displayed within the monitoring system

The RS gives post-hoc local interpretations to the suggestions given by the model. The interface should not only elaborate on the rationales behind the recommendations but also provide a thourough feedback to the user regarding their decision to accept or reject these recommendations.
The explanations provided for each aspect should be comprehensive and supported by external sources, particularly since the model influences significant financial decisions for households. Additionally, the UI should include a detailed overview of the model's training methodology, ensuring transparency and understanding.

Under this deployment strategy the model would be first trained once the reccomender system is first initiated by the user and then periodically retrained with new data to refine its predictions and adapt to changing market conditions.

7.3 User Interface (UI)

The demonstrative video of the UI can be seen under the following link: https://www.youtube.com/watch?v=CbG2Y-1GYr0.

The developed interface is divided into 6 stages:

0. Introduction Page (fig. 83) Provides the user with an overview of the tool, detailing its scope of use, advantages, and limitations. It also explains the underlying technology and the factors considered in generating its recommendations.

1. System Info (Appendix N). The interface necessitates information regarding the current setup, including consumption details, roof geometry, and the type of electricity tariff. This information can be entered manually by the user, imported from other home energy system design tools, or retrieved from the monitoring software of an existing system.

2. Recommendation Scope (fig. 84) - The user can define the scope of the optimization, the goal—whether to maximize both financial and environmental returns or to focus exclusively on one of these aspects—the

07 DEPLOYMENT

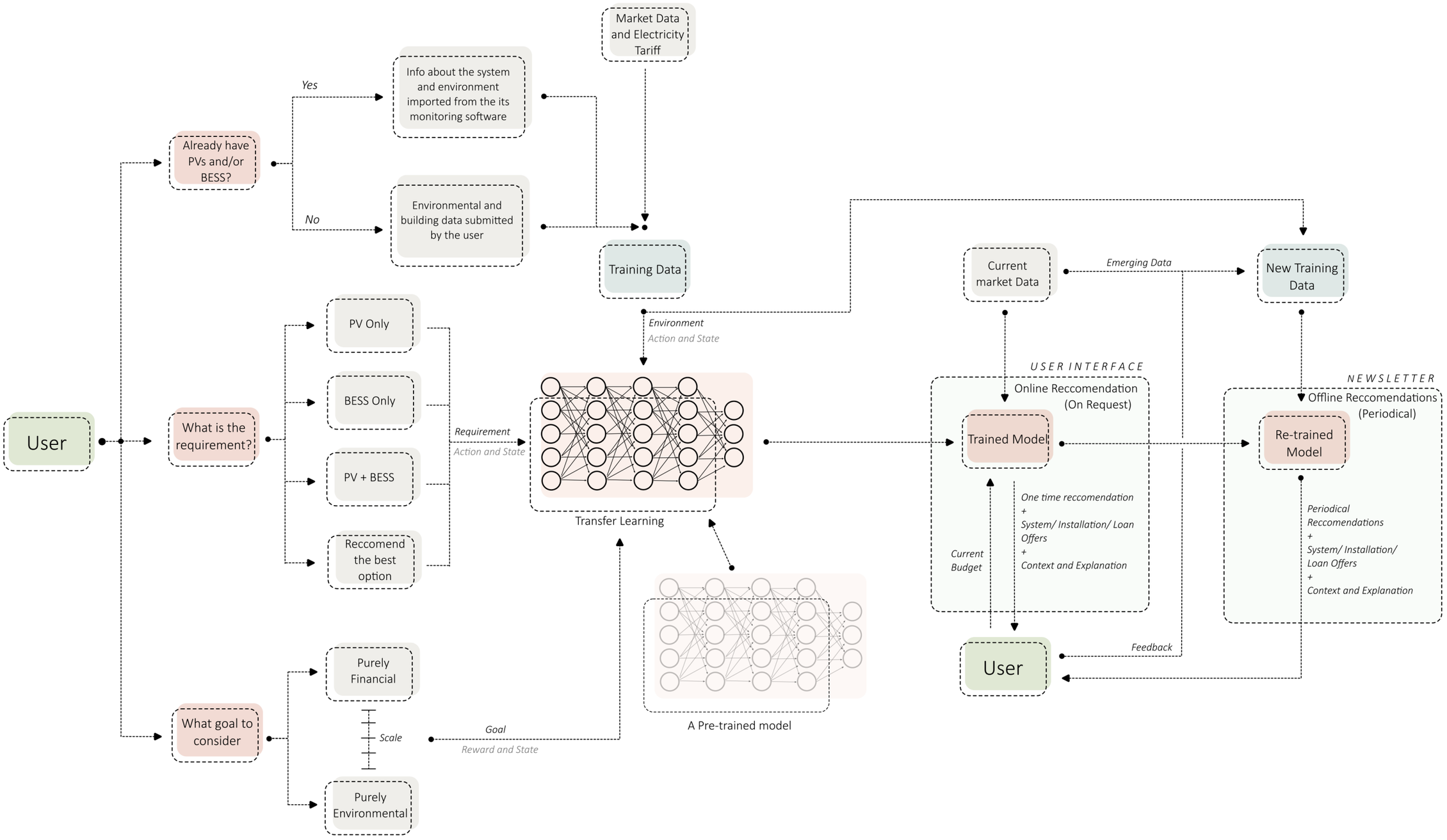


Figure 82: Model Deployment Flowchart

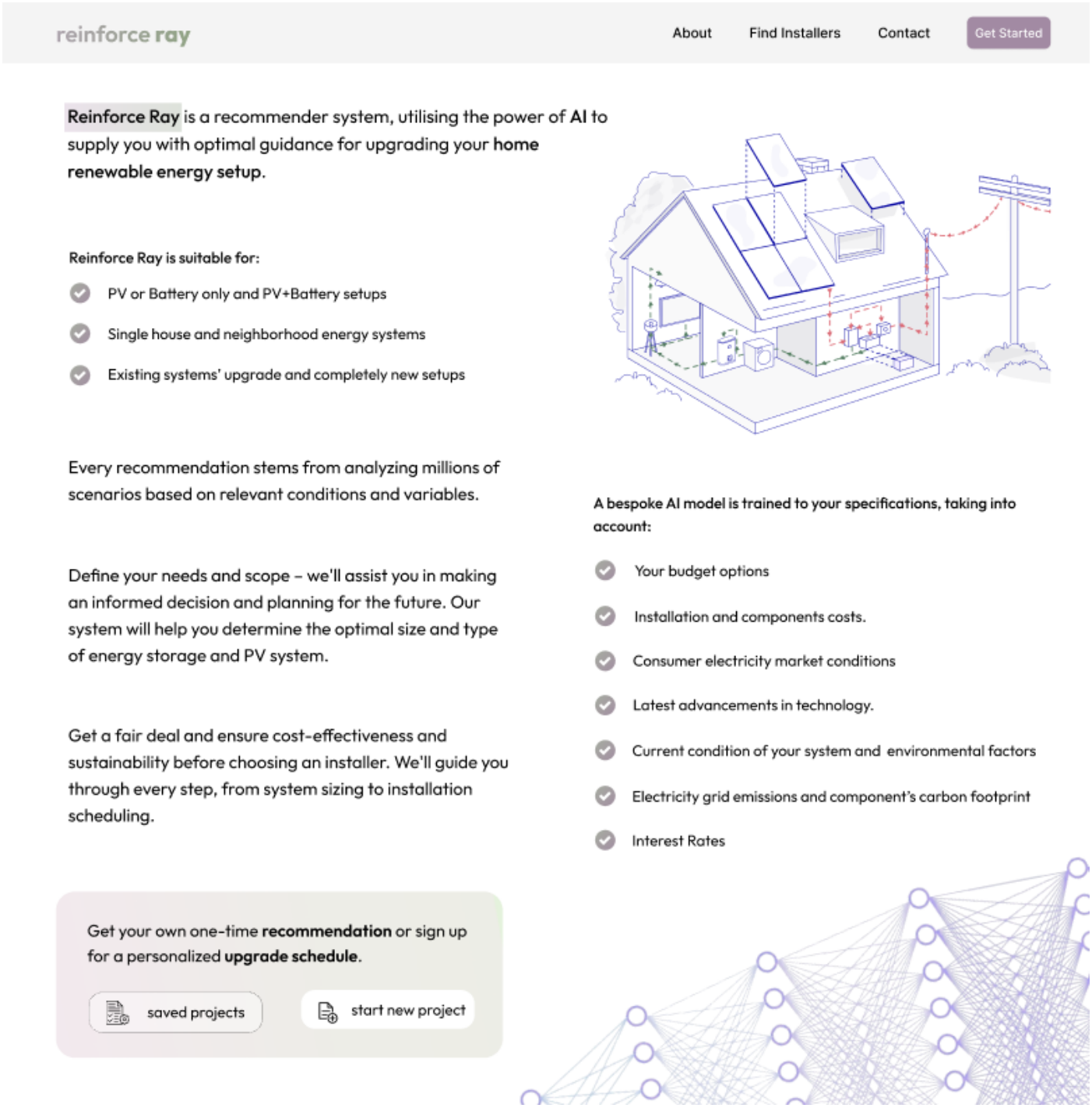


Figure 83: Reinforce Ray's UI Starting Page

maximum budget, and the recommended system type. Consequently, distinct models, each trained on different reward functions, are required to support this functionality within the UI.

3. Performance Overview (fig. 86, 87, Appendix N) – To enhance user comprehension of the forthcoming advice, the user interface presents an overview of the system's effectiveness, incorporating relevant data on its financial and environmental performance, as well as the contributions of each component.

4. Recommendation (fig. , Appendix N) – The recommendations are presented as the most confident option, alongside other viable alternatives (n-best alternative actions). In contrast to the deployment methods described in section 2.3.1, this is an online recommendation provided upon user request. Each advice includes relevant data on benefits and costs, as well as charts to effectively communicate the potential impact. Users can also see how varying budgets affect the model's output. Additionally, users can determine their eligibility for a loan to cover installation costs and browse best sale and loan offers for each recommendation.

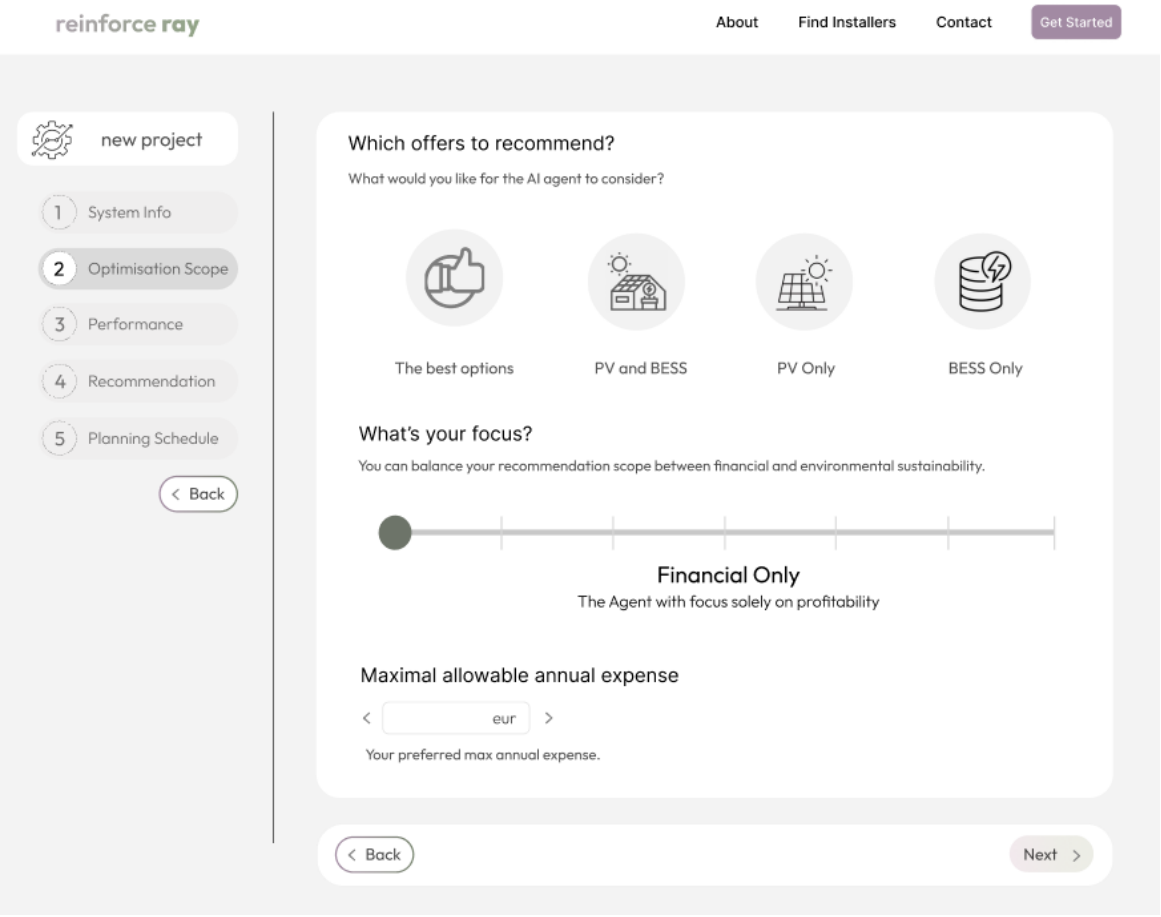


Figure 84: Reccomendation Scope Selection Page

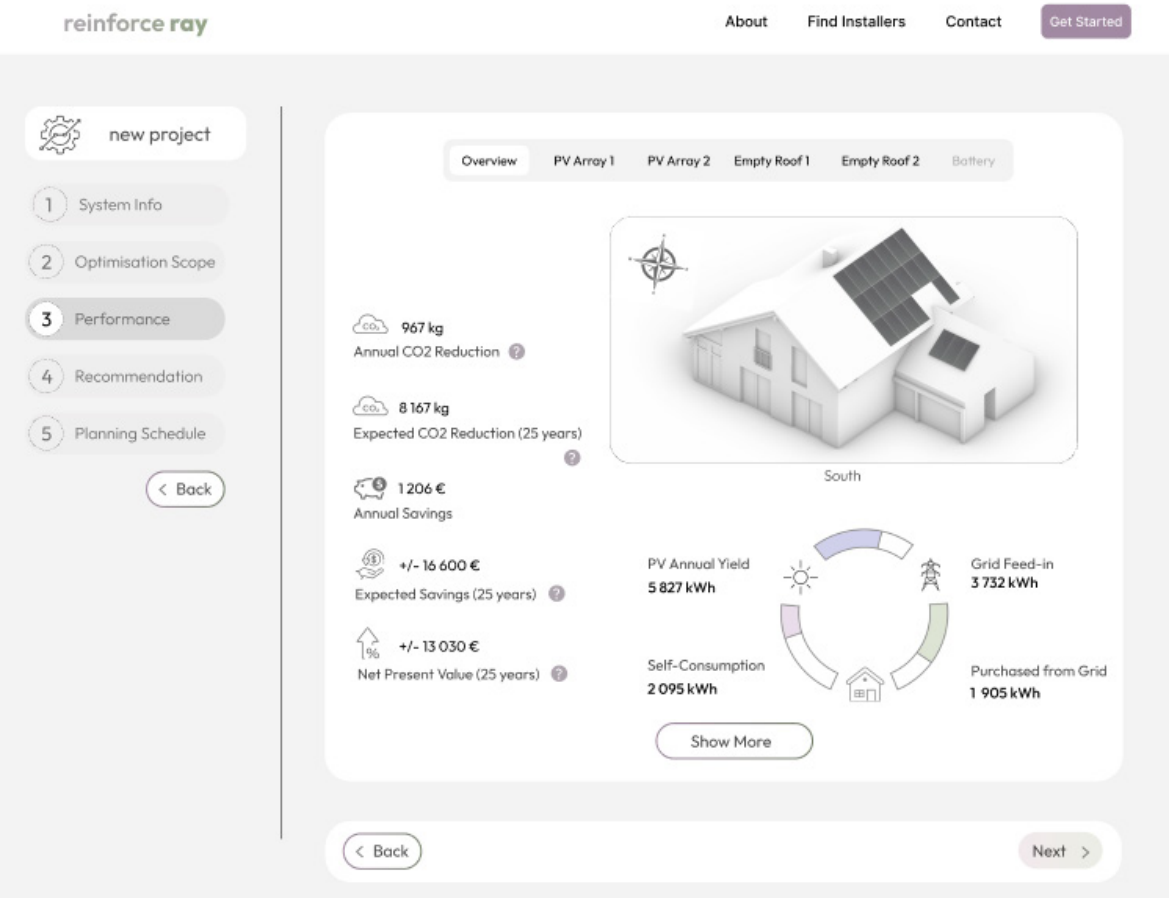


Figure 85: Performance overview page, other tabs of this page are shown in Appendix N

5. Planning Schedule (fig. 88 , Appendix N) To enable users to distribute costs over time and plan their expenses more precisely, they can enroll in a customizable upgrade plan. This process is executed in several sequential steps, with the system notifying users when to complete each one within the responsive strategy.

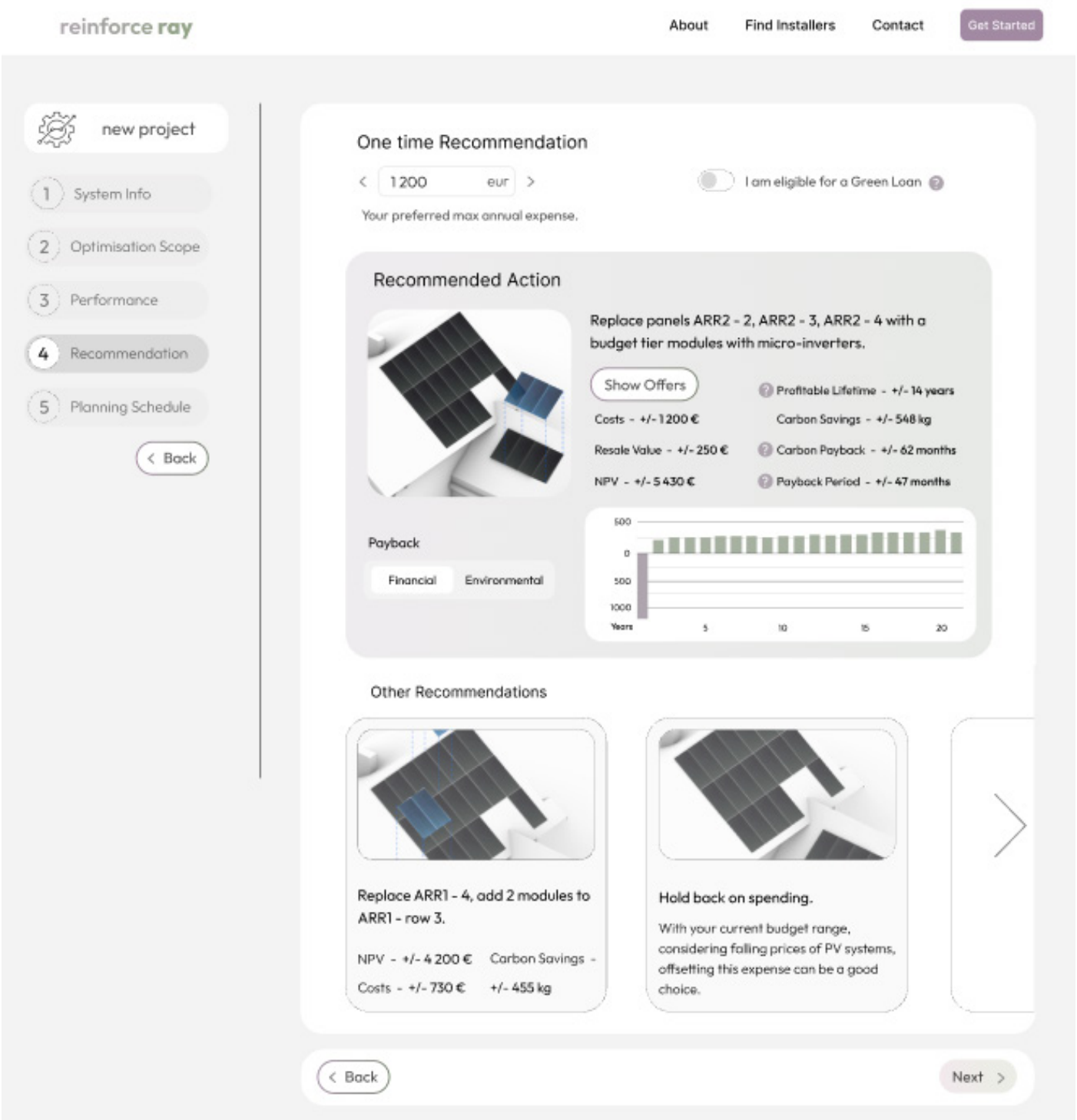


Figure 86: One-off recoemmendation for a budget-constraint, non loan eligible and PV only scope, secondary recommendations from the bottom of the figure, included in Appendix N

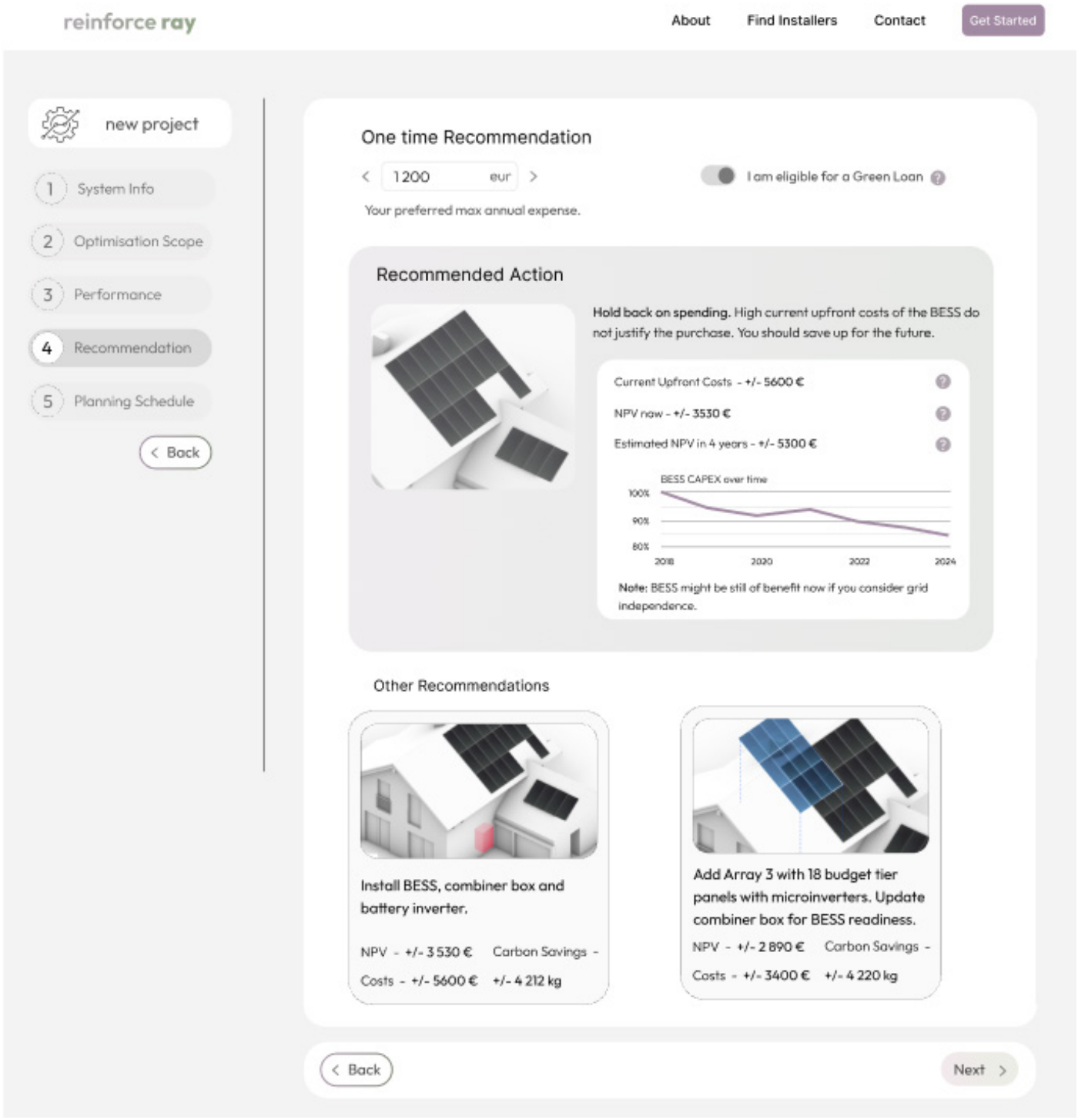


Figure 87: One-off recoemmendation for a budget-constraint, loan eligible and PV + BESS scope

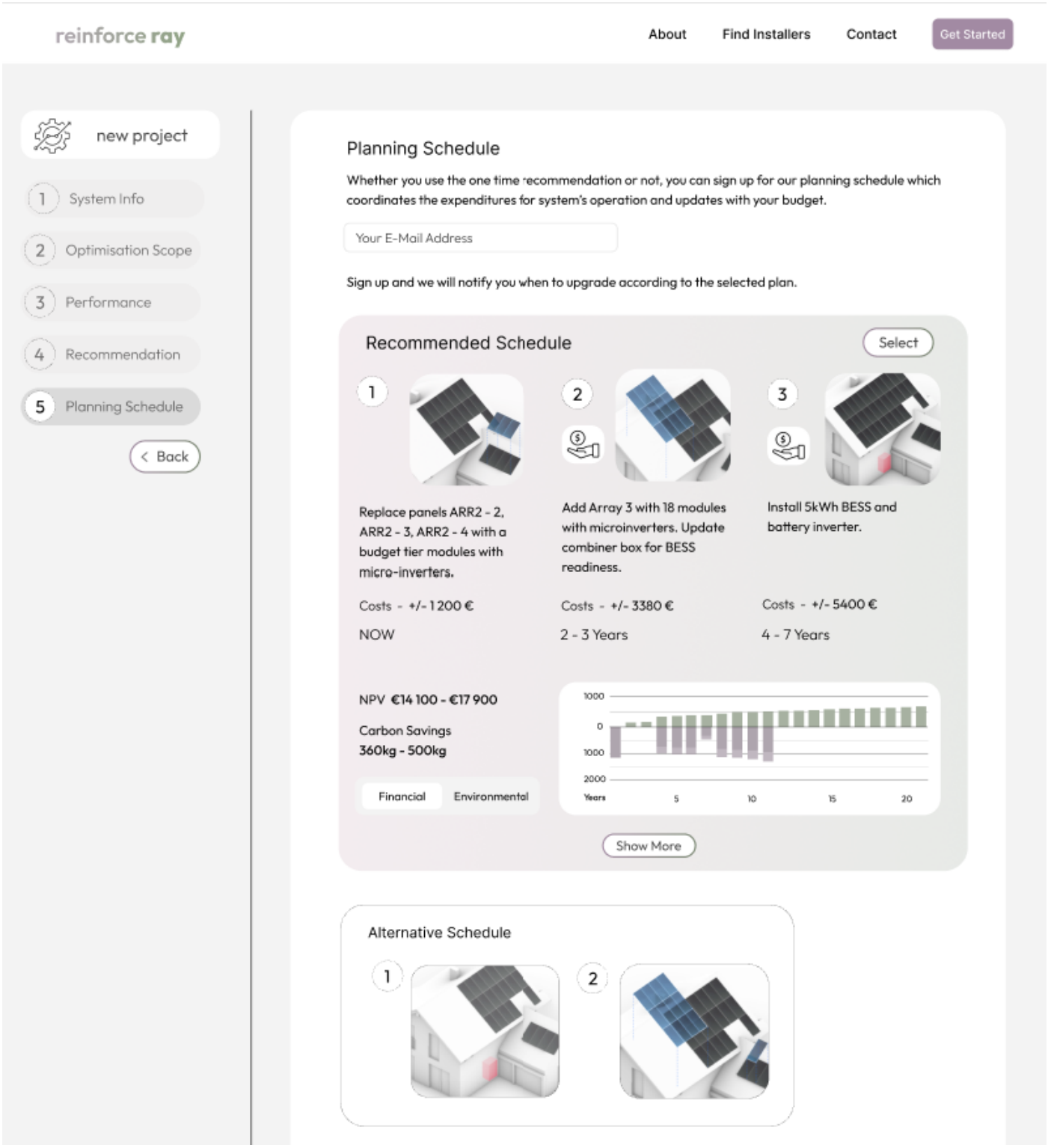


Figure 88: Planning Schedule page - sequential reccommendation presented to the user, continuation in Appendix N

08. Conclusion

08 CONCLUSION

8.1 Conclusion

This graduation project focused on developing a computational workflow using reinforcement learning for the optimal planning of residential PV and BESS systems. It aimed to navigate future uncertainties effectively by integrating and evaluating a trained RL model within a recommendation system. This system is designed to support stakeholders in making informed decisions regarding investments in solar power and energy storage.

The aim of this thesis was to answer the following research question:

How can reinforcement learning based recommendation workflow be used for long-term planning and design of residential grid-connected PV and BESS under the uncertainty of future scenarios?

In order to address the primary research question, several secondary questions were formulated, the answers to which are shortly refered to below.

How does the residential grid-tied photovoltaic system with BESS operate?

A residential grid-tied PV system with a BESS operates by generating electricity through solar panels during periods of sunlight. The solar panels, composed of photovoltaic cells typically made from various types of silicon, convert sunlight into DC electricity. This DC output is then converted into AC by an inverter, synchronizing it with the grid's voltage, frequency, and phase requirements. Excess electricity produced that is not immediately used by the household can be fed back into the utility grid. This process is facilitated by metering devices that record the amount of electricity redistributed to the grid.

Here, energy storage is often incorporated, predominantly using lithium-ion batteries due to their high energy density and longer lifespan. These systems can either be AC-coupled, involving multiple conversions between AC and DC which may lead to efficiency losses, or DC-coupled, which connects the solar panels directly to the batteries.

The setup sellected for the optimisation involves several different households, PV modules in different crystaline technologies and market ranges, microinverters and DC coupled BESS.

Which variables to include in the optimisation process?

The variables to include in this process should comprise both financial and environmental factors to evaluate their long-term economic efficiency and sustainability. Critical variables include:

- Energy Balance (building load profiles)
- Environmental Factors (solar radiation, shading, soiling, wind speed, ambient temperature, tilt angle and orientation)
- PV Related Factors (4 parameters, cabling and inverter losses)
- Financial Variables (PV module and BESS CAPEX (in eur/Wp), electricity tariff and its structure, ne-metering or feed-in-tariff, operational expenditures, replacement and installation costs, inverter and balance of system costs, resale revenue, inflation and loan interest rates)
- Environmental Variables (the grid emission factor and carbon enbodied in the materials, transport and manufacturing of the main systems and their secondary components)
- Improvement of the PV crystaline cell technology
- BESS and PV degradatrion and failure over time.

How can we evaluate the economic profitability and environmental benefits of rooftop PV systems and residential BESS?

The economic profitability and environmental benefits of rooftop PV systems and residential BESS can be evaluated using various optimization algorithms that focus on objective functions such as Net Present Value and Internal Rate of Return. NPV calculates the total value of the system, incorporating all electricity cost savings and subtracting the initial investment costs. This was used for the evaluation of the results and testing of the trained reinfircement learning models. This evaluation necessitates the inclusion of the relevant variables described in the previous section. in their complexity. To evaluate the environmental benefit the Net Carbon Savings was used.

What constitutes the most appropriate model for forecasting the electricity yield of photovoltaic systems?

The most appropriate model for forecasting the electricity yield of photovoltaic systems, as identified through comprehensive literature analysis, is the single-diode model. This model, characterized by its use of a four-parameter equivalent circuit, provides an accurate approximation of

08 CONCLUSION

PV performance under various conditions of irradiance and temperature. By modeling the electrical behavior of a photovoltaic cell through its current-voltage relationship and capturing key parameters such as light-generated current, diode saturation current, and series and shunt resistances, the single-diode model effectively reflects the intricate dynamics of a PV cell. This robust modeling approach is especially valuable for applications requiring precise electrical output predictions, thereby supporting informed decision-making in solar PV system installations.

What constitutes the most appropriate model for simulating the operation of a home BESS?

The most appropriate model for simulating the operation of a home BESS integrates various aspects including battery chemistry, operational management, and system degradation. Considering the complexities of battery operation, which involve factors such as depth of discharge, state of charge, power rating, and round-trip efficiency, a realistic simulation requires a nuanced approach. This selected model uses a linear approximation to predict the changes in the energy content of a lithium-ion battery, factoring in efficiency losses due to inverter operation, leakage, and charging and discharging processes. Additionally, a novel RL based approach was developed to realistically simulate the operation of the battery. This method was benchmarked against a simpler rule-based strategy to evaluate its effectiveness.

How to generate future scenarios of the identified optimisation variables for model training?

A plethora of probabilistic and stochastic processes was utilized to simulate all relevant variables. These approaches were selected based on their applicability to the use case, as well as extensive literature review, feasibility within the frame of this thesis project, data availability, testing, and experiments. The methodologies employed are grounded in probability theorems and stochastic differential calculus.

Which reinforcement learning algorithm and in what configuration is most suitable for this problem?

The field of RL was extensively explored, including various types and variants of RL algorithms and their applications. Based on criteria such as robustness and stability, scalability, multi-objective optimization capability, and learning from limited resources, two algorithms, DQN)and PPO, were initially selected for evaluation. However, initial testing revealed that DQN was incompatible with the multi-discrete action space used in the environment, leading to its replacement with Advantage Actor-Critic (A2C). Subsequent evaluations determined that PPO achieved better convergence than A2C. The hyperparameters and policy networks of PPO were finely tuned using both a specialised library and manual adjustments, and their performance was systematically evaluated.

What kind of action, observation spaces and reward function to consider?

RL training and evaluation environments were constructed using the OpenAI Gym framework. The observation (state) space was defined as a box space, consisting of a number of floating-point numbers that varied depending on the scenario. These included data pertaining to the state of each PV panel and the battery, identifying integers for PV panels, as well as current pricing information for system components, electricity, budget, interest rates, and expenditures.

The reward function for financial optimization incorporated relevant expenditures and profits derived from the system installation, including a penalty for exceeding the budget. Conversely, the reward function for environmental optimization factored in net carbon savings achieved by the system.

Two multi-discrete action spaces were assessed: one in which the agent made decisions separately for each panel and another where decisions were made for the entire system collectively. The latter option demonstrated superior performance.

How to deploy the trained model and how could it be used by the end user?

The proposed workflow involves creating a recommendation system, with a UI designed to provide real-time, optimal actions for PV system users. This system will collect data via an API on current energy prices and system components, and can operate both offline (providing periodic advice) and online (offering immediate recommendations). The system explains the reasoning behind its suggestions through the UI, assess their financial and environmental impacts, and allow users to customize their optimization preferences. To ensure relevance and accuracy, the model can be initially trained and periodically retrained with updated data, facilitating informed decisions in a dynamically changing market.

08 CONCLUSION

To answer the main research question an experimental workflow was formulated involving 4 phases:

1. Simplified toy problem, a testing and learning ground for the selected approach,

2. PV system optimisation encompassing all variables in their complexity,

3. PV and BESS optimization employing a methodology that builds on the previous step,

4. Demostrative deployment and UX formulation.

The computational workflow encompassed the integration of probabilistic and stochastic processes for simulating the uncertainties inherent in relevant variables, formulation of the problem as a MDP for sequential decision-making across 25 equal timesteps, and construction of a RL environment. This was followed by testing, evaluation, and refinement of the selected RL algorithm. The workflow was tested across a multitude of scenarios and variants, which included different environmental conditions, building loads, and types of technology involved. It was then evaluated against the base policy and various action space configurations. The plan for the trained and evaluated model integration was formulated into a user interactive test deployment.

Overall, it can be concluded that reinforcement learning **is a viable option for the planning and design of residential grid-connected PV and BES systems.** Furthermore, following the results from Chapters 4 and 5, it can be concluded that as the complexity of the environment increases—characterized by a greater number of optimization variables such as the number of panels, the presence of a BESS, and budgetary constraints—the performance of the model improves relative to the base policy. This finding presents a promising avenue for further research, which is discussed in the subsequent section.

8.2 Discussion

8.2.1 Alternative Computational Approaches

This thesis project proves that RL is a suitable approach for the problem at hand. Yet RL comes with several barriers and shortcomings: it requires extensive training and large amounts of data, which can be resource-intensive and time-consuming. Additionally, the complexity of designing appropriate reward functions and ensuring convergence to optimal policies can pose significant challenges. Furthermore, real-world deployment of RL models may face practical limitations, such as integration with existing systems and the need for ongoing maintenance and updates to adapt to changing conditions.

Traditional optimization techniques, such as Linear Programming and Mixed-Integer Linear Programming, are often used for municipal-scale planning of rooftop PV deployment. These methods rely on static models with fixed parameters, making them suitable for deterministic optimization. However, they may not adequately handle the uncertainties and dynamic changes inherent in long-term planning for renewable energy systems. Implementing LP or MILP for this problem would involve setting up a static model with predetermined parameters and constraints, which may not adapt well to long-term variations and uncertainties in the system.

Heuristic methods like Genetic Algorithms and Particle Swarm Optimization have also been applied to optimize PV and BESS systems. These methods explore a large solution space and can be effective for specific optimization problems. However, they might not balance exploration and exploitation as effectively as RL algorithms. They often require manual tuning and may not perform well in highly dynamic scenarios. Implementing GA or PSO would involve defining a fitness function to evaluate solutions, running multiple iterations to evolve the population of solutions, and selecting the best-performing ones

Model-based approaches, such as Dynamic Programming and Model Predictive Control, require accurate models of the environment to predict future states and optimize actions accordingly. This dependency on precise models can be a limitation when the environmental model is complex or not well-defined. Moreover, model-based methods can become computationally expensive as the state and action spaces grow. Implementing DP or MPC would necessitate developing precise models of the PV and BESS systems, forecasting future states, and solving optimization problems at each time step, which could be computationally intensive and less adaptable to changes.

8.2.2 Assumptions and Limitations

The observation (state) space within the einvironemnt are always fully observable. This means that at any time step, the agent can observe all the relevant variables and parameters that define the state. In contrast, in a partially observable state space, the agent merely receives observations that provide incomplete or noisy information about the true state, requiring it to infer the underlying state from these observations to make informed decisions. In practice, for example the current values of the BESS and PV degradation might not be explicitly available and supplied to the agent in its current forms.

The optimisation does not consider varying shading conditions or climate change impact on the solar radiation or cell temperature. The electricity production or degradation resulting from shading could be further implemented into the partially observable action space.

Analysed houses' consumption profiles are assumed to remain static over the optimisation period. However a more comprahensive optimisation could consider evolving consumption patterns or future retrofitting posibilites.

Furthermore, only the time-of-use electricity tariff type is taken into account. Ideally dynamic pricing should also be examined.

The utility grid carbon emissions are asumed to evolve according to the long-term decarbonisation plans set out by the relevant authorities in the Netherlands. Moreover, the simulations and stochastic scenario generations in this project have been conducted under the presupposition that climate change will not lead to extensive disruptions of the electricity grid and renewable energy supply chains.

Furthermore, it is assumed that the single-phase microinverters considered are optimally connected to each phase within a house at all times. Potential impact of the optimisation outcomes on the electricity grid are not taken into account.

8.3 Further Research Recomendations

The following section describes the possible paths for further research.

1. Benchmark Against Other Computational Approaches

This work is limited by its lack of comparative analysis essential to benchmark it against traditional optimization techniques like LP (Fan & Xia, 2017), MILP (Ren et al., 2023), GA, PSO, DP, MPC or a more sophisticated rule-based method under similar conditions.

2. Evaluate the Approach on Different Typologies

This thesis project is limited to single-family houses. It would be beneficial to evaluate the approach across various building typologies, such as **public buildings or multi-family residences**. Additionally, this approach could be further explored in a **solar microgrid scenario**, which consists of multiple prosumers with varying demands and production capacities (Khanal et al., 2023). This system based on energy sharing, could operate either independently or in conjunction with the main power grid.

3. Incorporate Other Objectives

Additionally, other objectives could be incorporated into the optimization framework, such as **grid independence** or coordination with the grid to avoid overloading it. Nuaturally, the environmental evaluation of the PV+BESS setup should be also conducted.

4. Incorporate Inverters into Optimisation

Inverters are not considered as an optimization variable in this study; however, they are a crucial component of the entire system. Different inverter configurations—such as string, micro, or DC-module inverters—along with their respective sizes, costs, and efficiencies, should be evaluated for their impact on the overall system performance.

5. Broaden the Optimisation Scope

Other devices integral to a home energy and heating system can be incorporated into the optimization, such as heat pumps, HVAC components, or high-energy-consuming household devices like refrigerators and other kitchen equipment. Additionally more PV specified equpiment can be considered such as presence of tracker systems.

More importantly, however, the optimisation should include other inverter options than the considered microinverters. Today still string inverters are most common technology on the residential market. Ideally, the model should be able to benchmark different array and inverter configuations against each other.

6. Make the Reccomendation Interactive

In its current state, the recommendation workflow does

not incorporate user feedback into the recommendations provided. The model could be enhanced to include a mechanism that learns and adapts the policy based on user responses to the recommendations. This adaptation could potentially improve the relevance and effectiveness of the system by aligning more closely with user preferences and requirements over time

7. Evaluate on a Real Scenario

As emphasized throughout this thesis, the developed approach was exclusively validated on synthetic data using the same methodology employed for training. It would be highly beneficial to validate it in a separate computational environment or through real-world implementation.

8. Test Different Algorythms

In this thesis projects three RL algorithms were tested and evaluated: A2C, DQN and PPO. Based on the available literature, however other algorithms might yeild better results and should be tested, ex Asynchronous Proximal Policy Optimization (APPO) or Soft Actor Critic (SAC).

Additionally **multi-agent RL** should be tested and evaluated, especially within the microgrid scenario with multiple prosumers at stake within the system.

9. Reevaluate Propabilities

All stochastic and probabilistic processes used for the simulation of scenarios across multiple variables should be reevaluated for their suitability. If necessary, they should be reconsidered and replaced with alternative methods.

10. Generalise across Different Environments

Currently, each use-case necessitates distinct training, as the models are specific to the environment, which encompasses factors such as the building's electricity load, roof shape, tilt, and orientation. Ideally the model included in the final tool should be exposed to extensive datasets, enabling it to generalize across different use-cases without requiring training for each individual case. The model could be trained on a dataset of household consumption patterns, or if one is not available a synthetic dataset generated using techniques such as bootstrapping and adding noise or generative adversarial networks (GANs). Alternatively the consumption profiles can be also clustered for different typology types using Markov Chains (MCs) combined with fuzzy c-means and Self-Organizing Maps (SOMs) (Morales et al., 2024).

8.4 Reflection

1. How is your graduation topic positioned within the Building Technology studio?

The MSc AUBS program, among other areas, emphasizes innovation in architectural engineering. It adopts a multidisciplinary approach, encouraging students to delve into topics that bridge various fields. This project is closely aligned with the Building Technology Master track as it concentrates on integrating renewable energy sources into buildings and employs advanced computational tools for multi-objective decision-making.

Photovoltaic technology combined with Battery Energy Storage Systems BESS represents the future of building-integrated energy systems, transforming homeowners into active participants within the electricity grid. This area of study falls under the Building Energy Epidemiology chair at the Architectural Engineering and Technology department, which focuses on developing methods for diagnosing, optimizing, and predicting the operational performance of energy systems.

Additionally, this thesis centers on developing an innovative computational tool that aids end-user decision-making. This development is rooted in cutting-edge computational technology, specifically reinforcement learning, a branch of artificial intelligence. Therefore, this project aligns well with the Design Informatics chair.

2. How did the research approach influence your design/recommendations?

The research initially established a robust theoretical foundation in areas such as machine learning, stochastic and probabilistic processes, photovoltaic and battery system modeling, and reinforcement learning-based recommender systems. The literature review effectively demonstrated how one research field naturally progressed to another. For instance, the modeling of photovoltaic panels and Battery Energy Storage Systems was significantly influenced by an initial analysis of residential grid-connected photovoltaic and battery systems, as well as the identification of factors influencing decision-making. These topics also contributed to the fosus of the examination of the literature about creating stochastic scenarios along with the study of reinforcement learning.

Following the literature review, an initial experimental approach was developed, tested, and evaluated through a stepwise iterative process. This method of research through experimentation provided valuable insights,

particularly concerning the main focus of the study: how reinforcement learning can be applied to the long-term planning and design of residential grid-connected PV and BESS in the face of uncertain future scenarios.

The detailed analysis of data, particularly the evaluation of model results and the capacity to identify opportunities within the data while systematically eliminating noise, are key strengths of the research methodology used in this study. This meticulous approach to data analysis ensured that the research findings were robust and of significant value.

3. How do you assess the chosen approach and methodology?

Each domain explored in this research presented unique challenges and complexities, yet all were crucial to the overarching goal of developing an informed computational decision-making tool for the adoption and development of sustainable home energy systems. To the best of the authors' knowledge, no similar study has been conducted to date, making it difficult to identify areas for potential improvement in the developed methodology. Despite covering advanced topics, that exceeded author's expertise in mathematics, computational science, and electrical engineering, access to numerous online tutorials and specialized Python libraries significantly facilitated the research; resources that were not available just a few years ago.

The research methodology, initially developed with an awareness of various obstacles and constraints—technical, logistical, temporal, and those related to the author's skills and experience—was initially deemed too conservative. It was later expanded to include BESS into optimisation.

While each explored area provided valuable insights, they also highlighted numerous opportunities for further research. This study would stronlgy benefit from benchmarking the reinforcement learning-based approach against other computational methodologies, such as linear or dynamic programming. Moreover, the feasibility of the optimized solutions should be evaluated in a separate, preferably real-life based environment.

4. How did the research approach work out and did it lead to the results you aimed for?

The findings of this research are promising, indicating that AI can effectively assist in decision-making for optimizing the type, sizing, and to a certain degree operation of home energy systems. Despite the complexity and interdisciplinary nature of this work, the results have surpassed the initial expectations. The primary objective of developing a computational workflow centered around a trained and evaluated reinforcement learning model was successfully achieved.

5. What moral or ethical issues did you encounter during the process?

With the rapid technological advancements in the fields of AI and machine learning over the past decade, particularly in generative AI, moral concerns have arisen regarding how AI might affect the roles of architects and engineers in the future. This thesis project's potential implementation could conflict with the interests of specialists, engineers, and businesses involved in home microgrid system design.

Furthermore, the anonymity of the building electricity loads dataset presented challenges by limiting the available information for the study. Notably, the dataset excluded details such as roof shape, orientation, or tilt to protect residents' privacy. To address this issue, these parameters were simulated based on the study of typical configurations and to test the developed approach across a variety of environments.

6. What is the impact of your project on sustainable development?

Sustainability was the primary motivator for initiating this project. As government institutions across Europe, both national and transnational, strive to decarbonize electricity grids by setting short-term and long-term carbon emission targets, this thesis could become a component of the broader strategy in planning and assessing the feasibility of renewable energy sources, such as PV and BESS, on local scales. This includes integration into home energy systems and microgrids, assisting stakeholders in transitioning from grid consumers to active prosumers who manage their own energy sources.

7. What is the socio-cultural and ethical impact?

One of the main goals of this thesis was to empower various stakeholders to make more informed decision about their building integrated energy systems. On the societal front, the research directly addresses the need for more efficient and environmentally friendly energy solutions at the household level. The development of a robust RL-based framework provides homeowners with

a clear strategy to optimize their energy costs and reduce their carbon footprint. The outcomes of this research could potentially lower the barriers to the adoption of PV and BESS by demystifying the economic and ecological trade-offs involved and by establishing a planning schedule. Furthermore, this project contributes to overall risk mitigation by enhancing the reliability and resilience of energy systems in residential settings.

8. How do you assess the academic value of your graduation?

From a scientific perspective, this project contributes to the field of applied energy informatics, home energy management systems, building energy epidemiology and design informatics by demonstrating how advanced computational techniques can solve real-world energy management challenges. It explores dynamic optimization and real-options approaches, which consider uncertainties in long-term energy planning.

9. How does this approach relate to the larger context of innovation within the broader industry?

Recent developments in natural language-based machine learning models, such as ChatGPT, offer potential synergies for decision-making tools. Traditionally, these tools provide static, text-based information. A more collaborative, interactive approach could allow decision-makers to query solutions and engage actively with the tool, enhancing their learning process. This shift might also redefine the role of experts as facilitators in decision-making. Moreover, integrating these tools with a broader network of data-driven insights in the built environment, such as the extensive geospatial data available in the Netherlands, could streamline the use of precise and detailed building data in sustainability decisions.

Moreover, as computational tools increasingly integrate into the built environment, particularly in decision-making and design assistance, tools like Grasshopper are utilized for climate and structural design optimizations. Adopting AI-based approaches could represent an advancement over current parametric methods.

Finally, the trained model can be effectively used in industry for strategic investment decisions PV and BESS. The integration of real options planning with this computational approach allows for the valuation of flexibility in investment decisions, acknowledging the right but not the obligation to pursue or defer certain investments based on evolving conditions. This methodology not only quantifies the financial benefits of strategic flexibility but also enhances the capability to manage risk more effectively in the face of uncertainty.

10. How do you assess the value of the transferability of your project results?

The concept of transferability and scalability is crucial for this thesis topic. This research aimed to create a scalable decision-making tool that assists homeowners and other relevant stakeholders in transitioning to sustainable electricity production and storage. The tool, developed on top of a computational workflow, utilizes the existing system monitoring setup integrated into the PV + BESS system. It is designed to suggest the most optimal actions to users sequentially, providing the necessary financial and environmental context for a given recommendation along with the best market offers that correspond with it.

11. What are the potential challenges limiting the use of the applied methodology in practice?

Currently, each use-case necessitates distinct training, as the models are specific to the environment, which encompasses factors such as the building's electricity load, roof shape, tilt, and orientation. Ideally, a pre-trained RL model would be incorporated into the final tool. This model would be trained on extensive datasets, enabling it to generalize across different use-cases without requiring training for each individual case.

Additionally, in its current state, the recommendation workflow does not incorporate user feedback into the recommendations provided. The model could be enhanced to include a mechanism that learns and adapts the policy based on user responses to the recommendations. This adaptation could potentially improve the relevance and effectiveness of the system by aligning more closely with user preferences and requirements over time.

Another notable limitation is that the system mandates the use of a specific microinverter technology, which is less prevalent in the market compared to conventional string inverters. Another limitation is that the optimisation is configured under the time-of-use electricity tariff only.

09.

Appendix A
References
List of Figures
List of Tables

APPENDIX A

I_L is approximated by the short-circuit current ($I_{L,STC}$) provided by the manufacturer under STC obtained from the following equation (Shongwe and Hanif, 2015):

$$I_L = G \times \left(I_{L,STC} + K_i \Delta T\right)$$

where:
G is the ratio of the irradiance with respect to STC value equal to 1 kW/m2,
K_i is the temperature coefficient for current,
ΔT is the temperature difference $T_{ambient} - T_{c,STC}$,

I_0 requires knowledge of the cell temperature and the material properties of the PV cell. It can be calculated as (Theristis et al., 2018, pp. 671–706):

$$I_0 = C_s \times T_C^{\,3} \times \frac{-E_g}{e^{n \times k \times T}}$$

where:
C_s is the material specific constant that includes the properties of a semiconductor
T_C is absolute temperature of the PV cell in Kelvin,
E_G is the bandgap energy of the semiconductor material used in the PV cell.

R_s can be estimated based on based on the module's open-circuit voltage (V_{mp}) and short-circuit current (I_L). R_{sh} is calculated from the I-V curve at very low currents. The equations for calculating R_s and R_{sh} can be found in Shongwe and Hanif p. 940 (2015).

n must be determined through testing, common assumption is between 1 and 2.

V_{th} is given by the formula:

$$V_{th} = \frac{k \times T}{q}$$

k is the Boltzmann constant (1.380649×10^{-23} J/K),
T is the absolute temperature of the p-n junction in Kelvin (K),
q is the elementary charge (1.602176634×10^{-19} C)

APPENDIX B

For the toy problem, the average hourly electricity yield of the house, as documented by Wahi (2023), was utilized. This data was then manually adjusted to align with the daily average normalized fluctuations of household electricity demand for the typical Dutch household, according to the study by Klaassen, Frunt, and Slootweg (2015). It is shown in figure 31.

APPENDIX C

The historical prices per kilowatt-hour (eur/kWh) have been compiled from a dataset representing a single-family household in Tilburg, covering the period from 1985 to 2018 (Verhoeven, 2022). For the years 2019 to 2023, the data has been supplemented with the average annual consumer energy tariffs as published by the National Service for Entrepreneurship, Netherlands (Rijksdienst voor Ondernemend Nederland, 2024)(fig. xx).

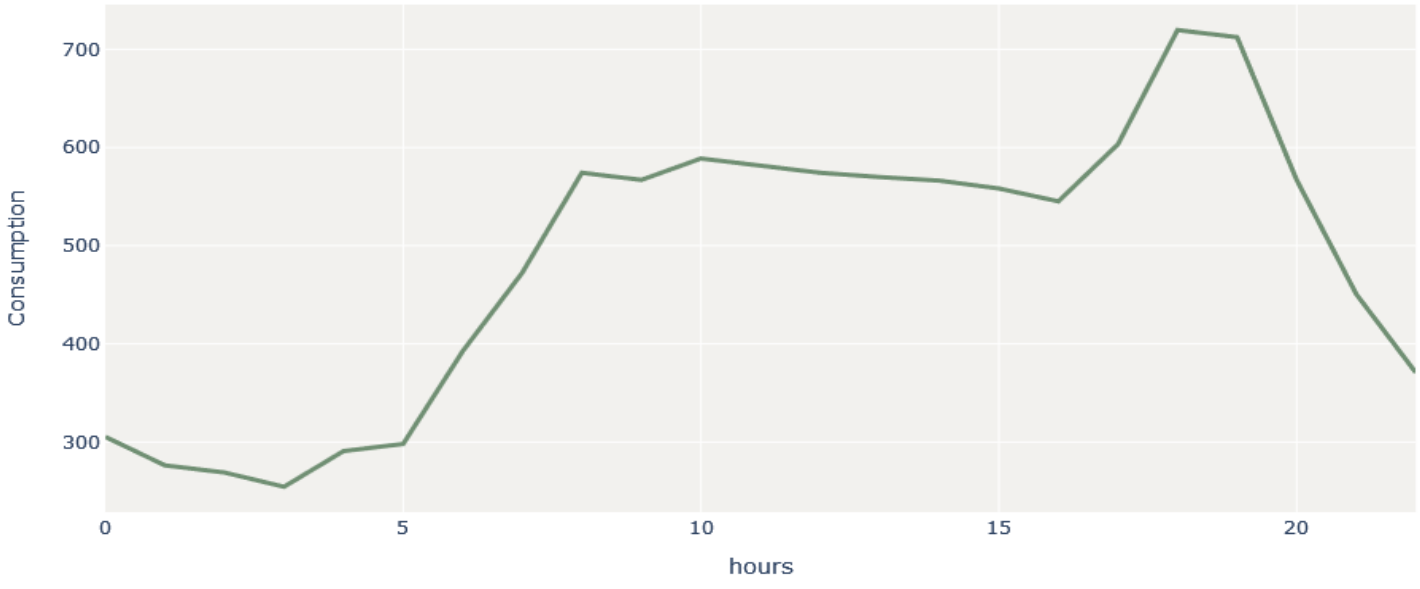


Figure 89: Daily Concumption profile used for the toy problem

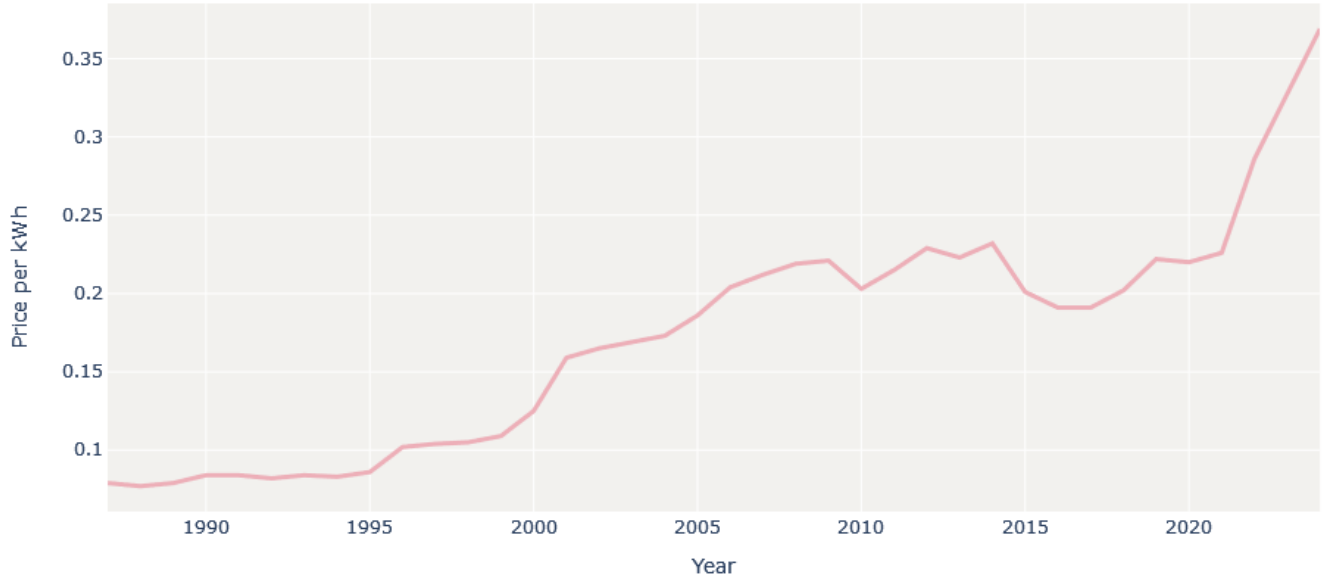


Figure 90: Historial Annual Electricity Tariffs

APPENDIX D

The historical emissions (kg CO_2 eq.) per kWh from 2015 to 2023 (Statistics Netherlands, 2023) are compiled with predicted emissions based on the Dutch Environmental decarbonisation goals until 2050 (Enerdata, 2024) (fig. 34).

APPENDIX E

The dataset is based from PV power plants based in the US. Figure xx depicts the mentioned temperature zones.

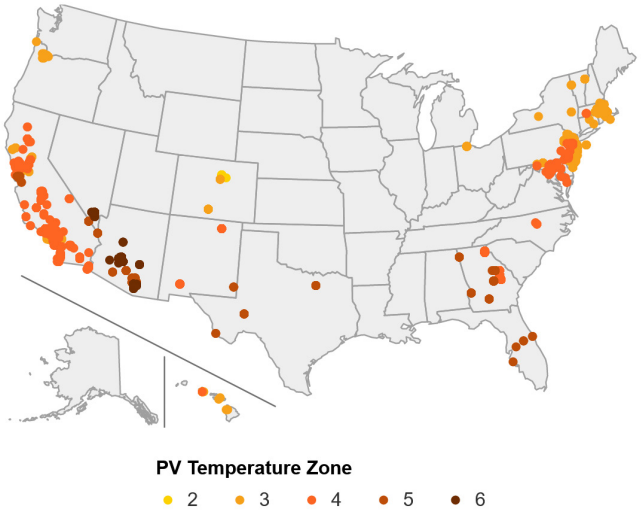


Figure 93: US PV temperature zones fir each sustem in the database (National Renewable Energy Laboratory, 2022)

APPENDIX F

Literature claims that monocrystalline solar panels are expected to last 30 to 35 years (Kumar and Sarkar, 2013), hence 30 is assumed for the budget panel. On the other hand, SunPower claims that 70% their panels stay operational in the 40th year since installation. Hence the parameter phi is conservatively assumed as 50. Furthermore for the used polycristaline panel, the longevity of which rarely exceeds 25 years (Gyamfi et al., 2023), asuming that it has been exploited in normal conditions for several years, the remaining lifetime is set to 15. The reliability of the microinverters is not taken into account, literature claims that their survival usually far exceeds the 25 year period considered in this study (Afridi et al., 2023).

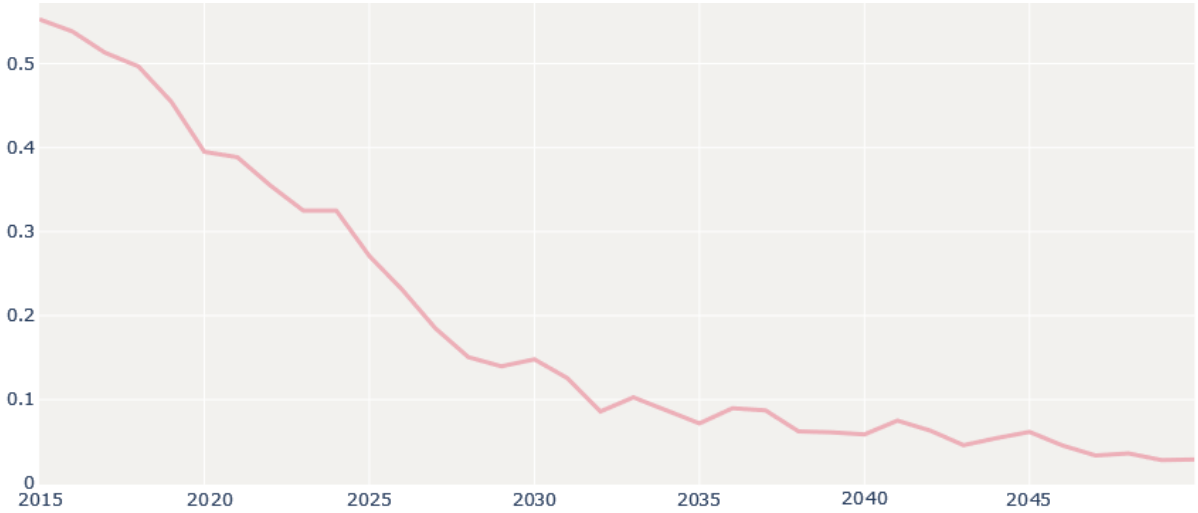


Figure 92: kg CO2 eq. per Wh of Electricity in the Netherlands

APPENDIX G

P_{max}	A	N_s	I_{sc}	V_{oc}	I_{mp}	V_{mp}	α_{sc}	β_{oc}	a_{ref}	I_L	I_o	R_s	R_{sh}	γ_r
436 Wp	1.87 m^2	66	11.57 A	48.2 V	10.82 A	40.3 V	0.0048	-0.113	1.642	11.59	$2.03e^{-11}$	0.245	138.4	-0.276

P_{max}	A	N_s	I_{sc}	V_{oc}	I_{mp}	V_{mp}	α_{sc}	β_{oc}	a_{ref}	I_L	I_o	R_s	R_{sh}	γ_r
265 Wp	1.6 m^2	60	9.07 A	38.3 V	8.39A	31.6 V	0.0045	-0.126	1.593	9.085	$3.22e^{-10}$	0.229	130.1	-0.43

P_{max}	A	N_s	I_{sc}	V_{oc}	I_{mp}	V_{mp}	α_{sc}	β_{oc}	a_{ref}	I_L	I_o	R_s	R_{sh}	γ_r
415 Wp	1.87 m^2	54	14.02 A	37.5 V	13.13 A	31.6 V	0.0054	-0.097	1.375	14.03	$1.97e^{-11}$	0.119	104.3	-0.333

V_{ac}	P_{so}	P_{dco}	P_{aco}	V_{dco}	C0	C1	C2	C3	P_{nt}	$V_{dc(max)}$	$I_{dc(max)}$
240 V	1.3	349	322.9	60	-0.0...27	-0.0...25	-0.0029	-0.0377	0.021	43	5.3828

APPENDIX H

$$LoanTerm(t) = \begin{cases} 1, & if\ \frac{C(t)}{B(t)} < 2 \\ 2, & if\ 2 \leq \frac{C(t)}{B(t)} < 3 \\ 3, & if\ 3 \leq \frac{C(t)}{B(t)} < 4 \\ 4, & if\ 4 \geq \frac{C(t)}{B(t)} \end{cases} \qquad eq.\ xx$$

where:
$C(t)$ is the annual expenditure
$B(t)$ is the annual budget
$LoanTerm$ is the repayment period in years.

The interest rate is assumed remain constant throughout the repayment period.

APPENDIX I

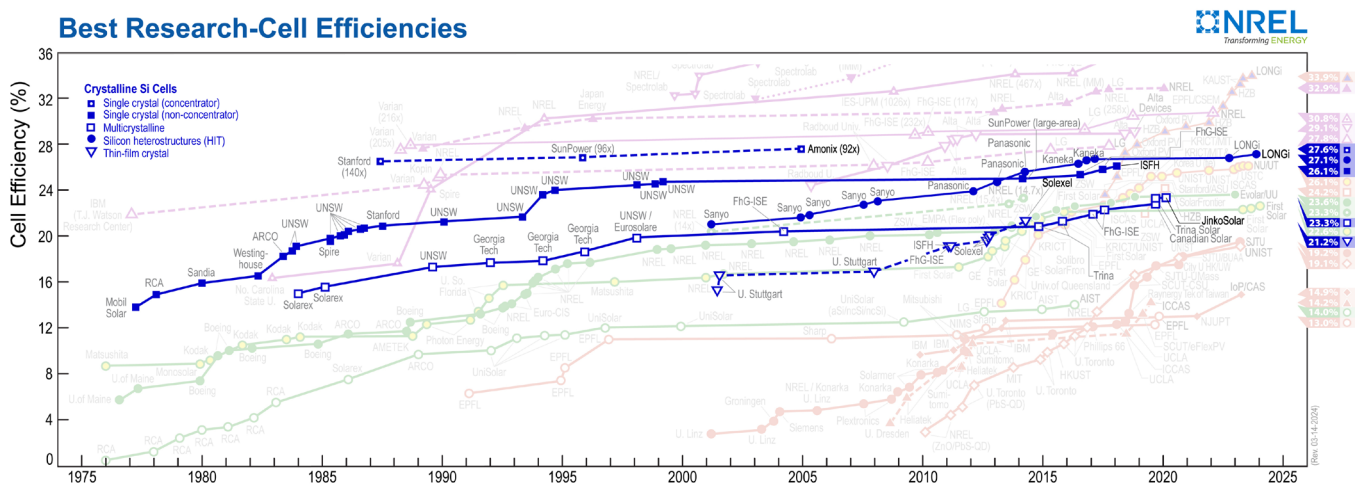


Figure 94: Crystalline silicon best research-cell efficiency chart

APPENDIX J

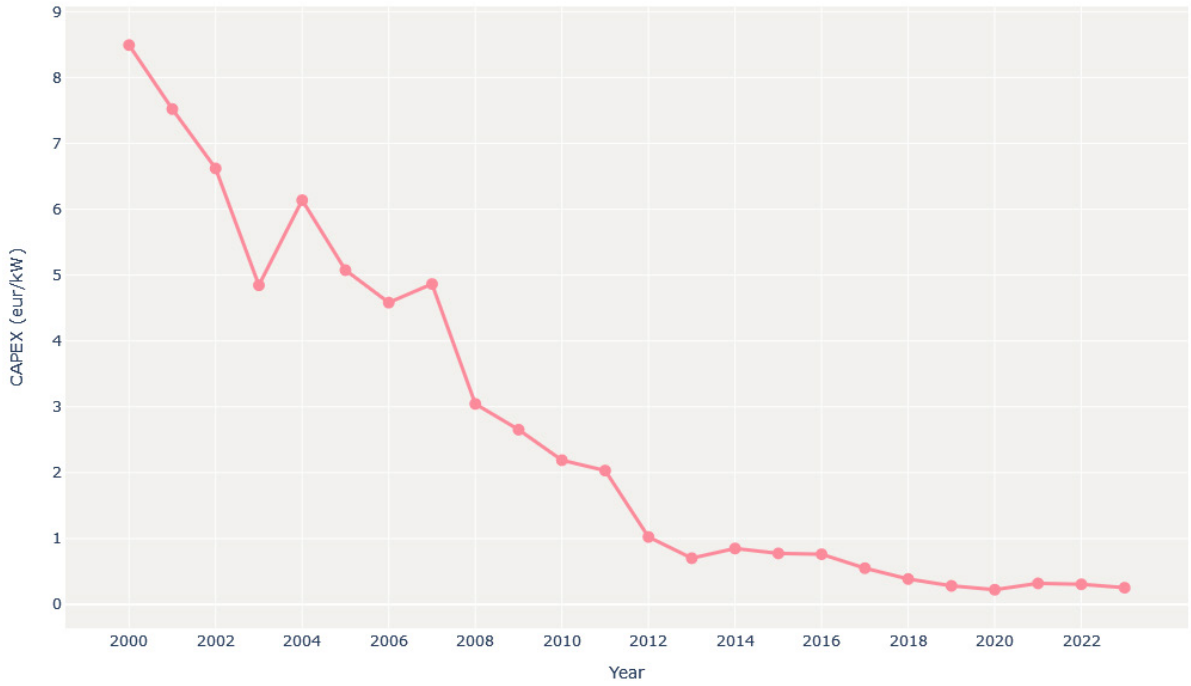


Figure 95: CAPEX (eur/kW) of a photovoltaic module on the German market (IRENA, 2022) (Jäger-Waldau , 2016)

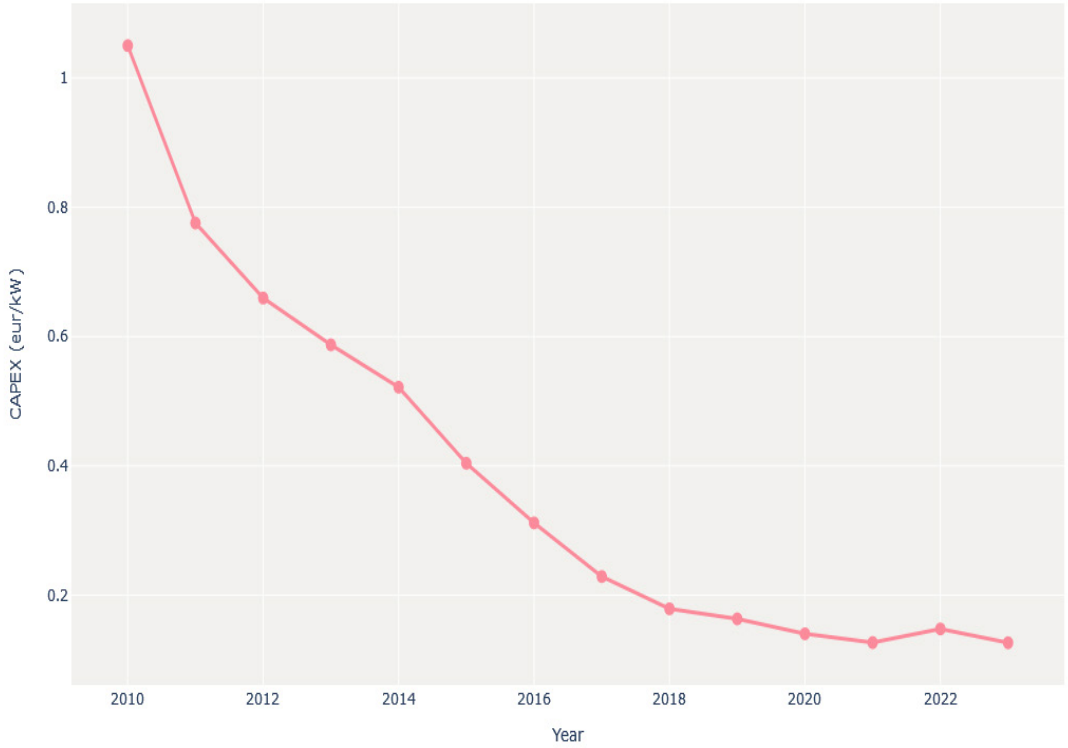
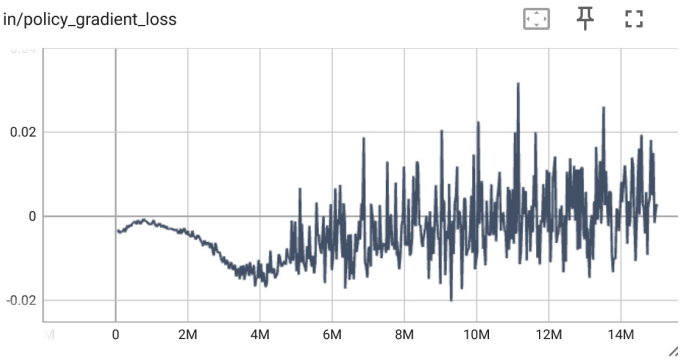
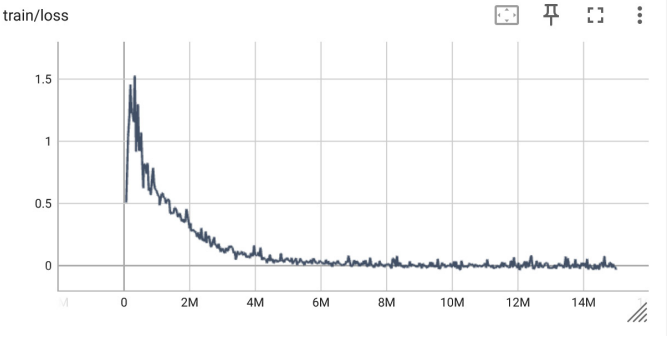


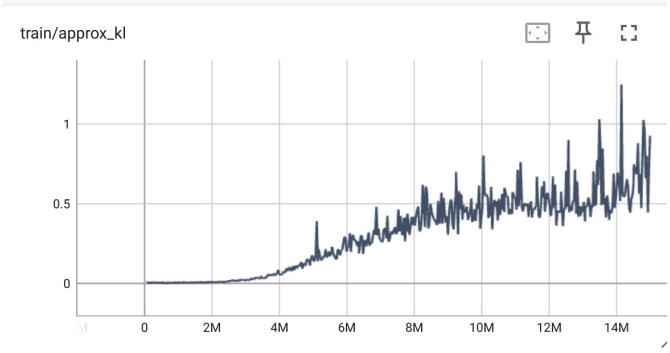
Figure 96: CAPEX evolution of lithium-ion energy storages

No of Neurons *Batch Size* *Lr Schedule* *Target Lr*	- Budget panel, - low budget, - house 2, - simple action	- Select panel, - high budget, - house 1, - complex action
Base Policy	20 695	8 190
256 256 Flat -	17 652	4 831
256 256 Linear 0.00001	17 882	4 874
256 512 Linear 0.00001	18 083	5 009
256 1024 Linear 0.00001	18 219	5 003
512 256 Linear 0.00001	20 109	6 641
512 512 Flat -	20 816	6 967
512 1024 Flat -	21 006	7 018
512 1024 Linear 0.00005	21 041	6 953
512 2048 Flat -	20 726	6 581
512 1024 Exponential 0.00001	22 102	7 542
512 512 Exponential 0.00001	22 057	7 305
512 512 Linear 0.00001	22 048	7 296
1024 256 Linear 0.00001	21 947	6 684
1024 512 Linear 0.00001	22 212	7 461
1024 1024 Linear 0.00001	22 228	7 559
1024 2048 Linear 0.00001	21 971	7 392
1024 512 Exponential 0.00001	22 213	7 481
1024 1024 Exponential 0.0001	22 231	7 516
1024 2048 Exponential 0.00001	21 982	7 370
1024 1024 Flat -	21 909	7 492

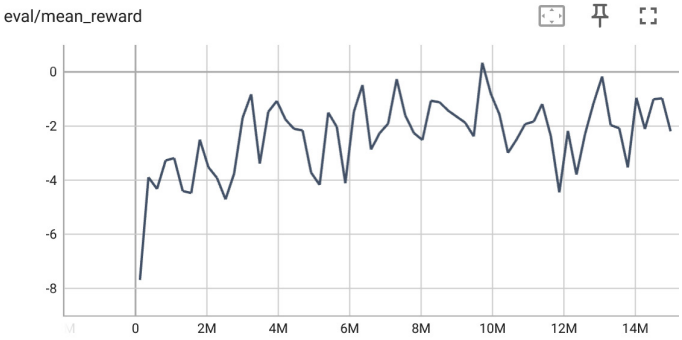
2048 1024 Linear 0.00001	22 010	7 586
2048 2048 Linear 0.00001	21 910	7 399
2048 1024 Flat -	22 109	7 562
2048 2048 Flat -	22 091	7 421
2048 512 Flat -	20 812	6 888

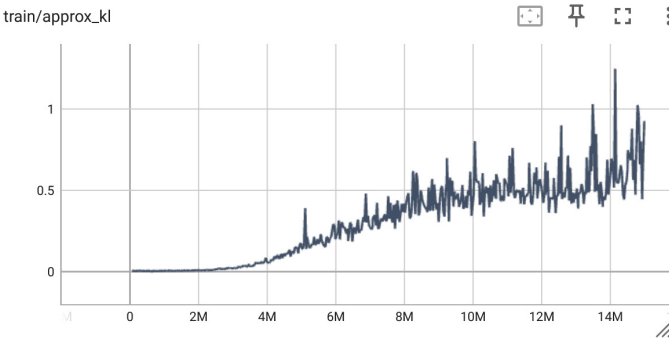


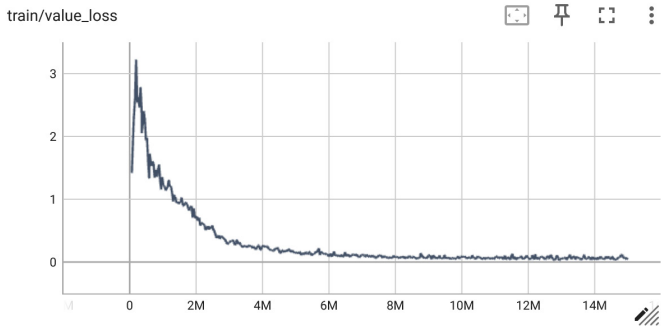




APPENDIX L







APPENDIX M

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	80	-	17	62	-	0
Budget Panel	186	-	104	115	-	49
Hi-End Panel	588	-	247	401	-	210
Select Panel	202	511	157	172	322	98

Table 21: B) Evaluation Results for House 1 - mean total Interests in Euro per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	80 703	-	76 425	80 703	-	80 411
Budget Panel	125 431	-	119 277	125 431	-	124 800
Hi-End Panel	133 492	-	124 145	133 492	-	127 552
Select Panel	124 836	125 585	120 049	124 836	126 084	122 090

Table 21: C) Evaluation Results for House 1 - mean total energy produced in kWh per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	213	-	44	114	-	41
Budget Panel	434	-	218	203	-	114
Hi-End Panel	993	-	404	621	-	422
Select Panel	471	876	352	315	489	208

Table 22: B) Evaluation Results for House 2 - mean total Interests in Euro per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	152 336	-	149 004	152 336	-	151 593
Budget Panel	236 617	-	224 419	236 617	-	230 335
Hi-End Panel	248 724	-	225 902	248 724	-	232 069
Select Panel	233 281	204 954	221 284	233 281	212 285	227 116

Table 22: C) Evaluation Results for House 2 - mean total energy produced in kWh per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	639	-	265	490	-	151
Budget Panel	689	-	392	633	-	261
Hi-End Panel	1 858	-	856	1 486	-	512
Select Panel	965	1 411	554	781	794	331

Table 27: B) Evaluation Results for PV+BESS House 2 - mean total Interests in Euro per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	152 336	-	147 298	152 336	-	151 388
Budget Panel	236 617	-	219 301	236 617	-	225 300
Hi-End Panel	248 724	-	223 902	248 724	-	227 326
Select Panel	233 281	235 380	216 623	233 281	214 620	226 864

Table 27: C) Evaluation Results for PV+BESS House 2 - mean total energy produced in kWh per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	520	-	113	503	-	84
Budget Panel	578	-	261	410	-	165
Hi-End Panel	1 427	-	646	1 156	-	277
Select Panel	694	496	411	437	322	264

Table 26: B) Evaluation Results for PV+BESS House 1 - mean total Interests in Euro per episode

	Low Budget			High Budget		
	Base Policy	*Model*		*Base Policy*	*Model*	
Action Space	-	*Complex*	*Simplified*	-	*Complex*	*Simplified*
Second-Hand Panel	80 703	-	75 896	80 703	-	78 616
Budget Panel	125 431	-	119 758	125 431	-	123 562
Hi-End Panel	133 492	-	127 040	133 492	-	128 755
Select Panel	124 836	124 117	118 868	124 836	124 176	120 789

Table 26: C) Evaluation Results for PV+BESS House 1 - mean total energy produced in kWh per episode

APPENDIX N

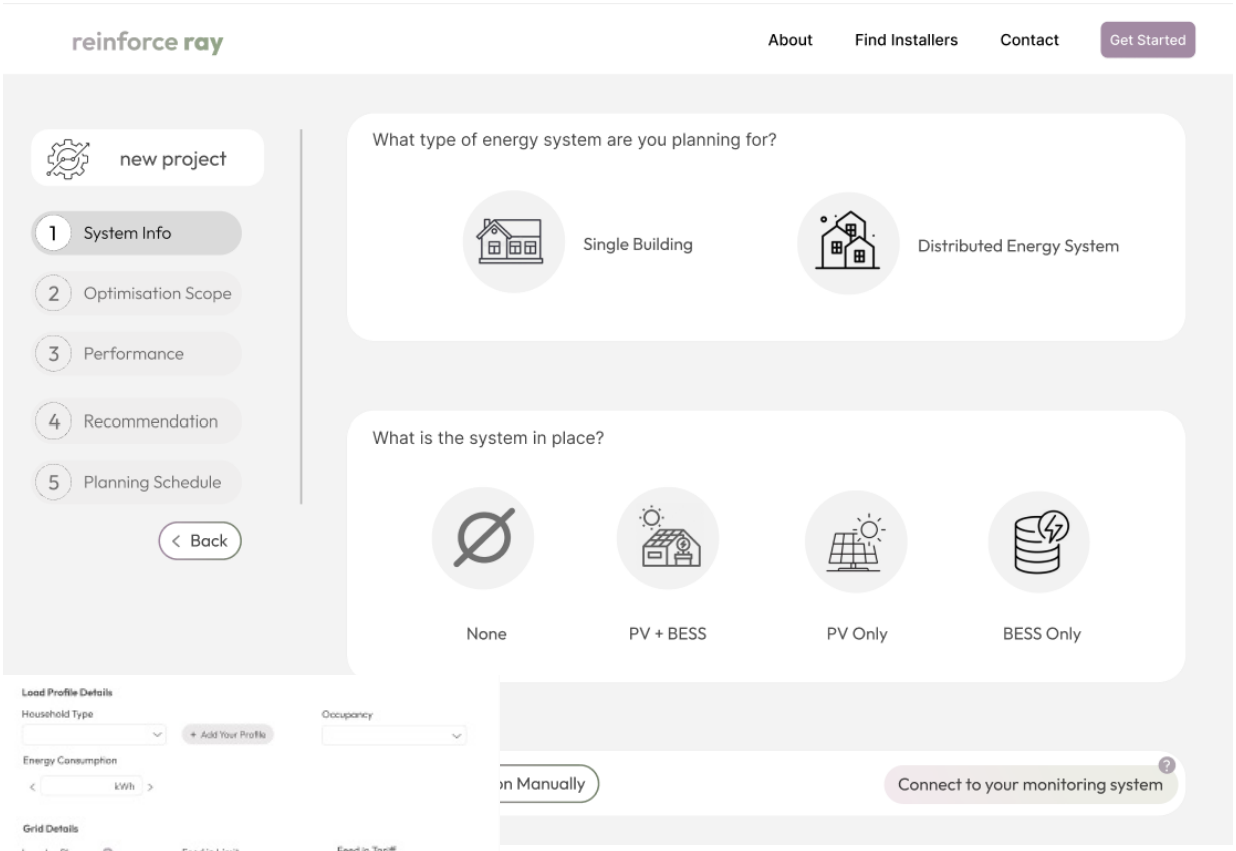


Figure 97: System Info page of the UI

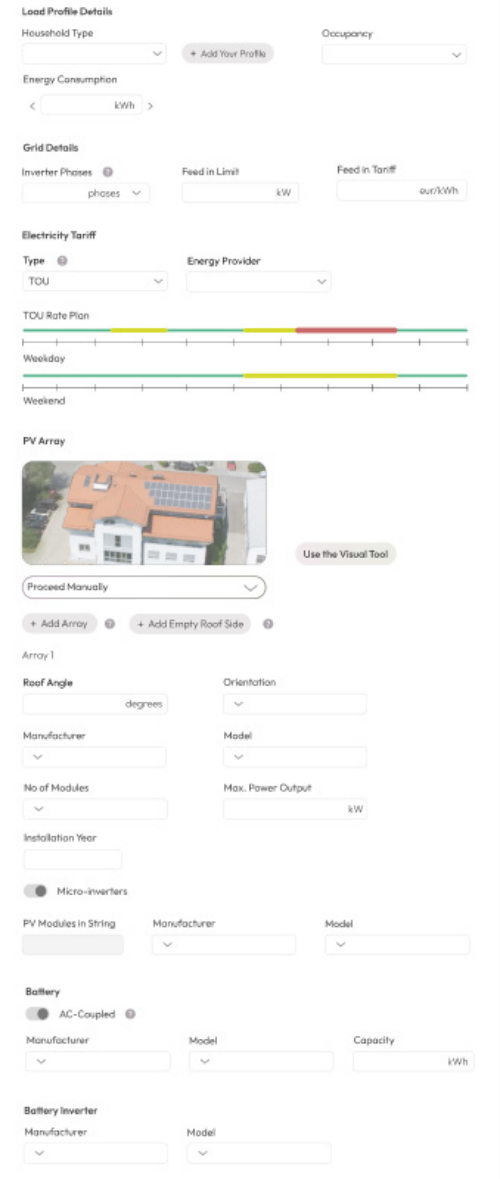


Figure 98: Manual System info input pop-up screen

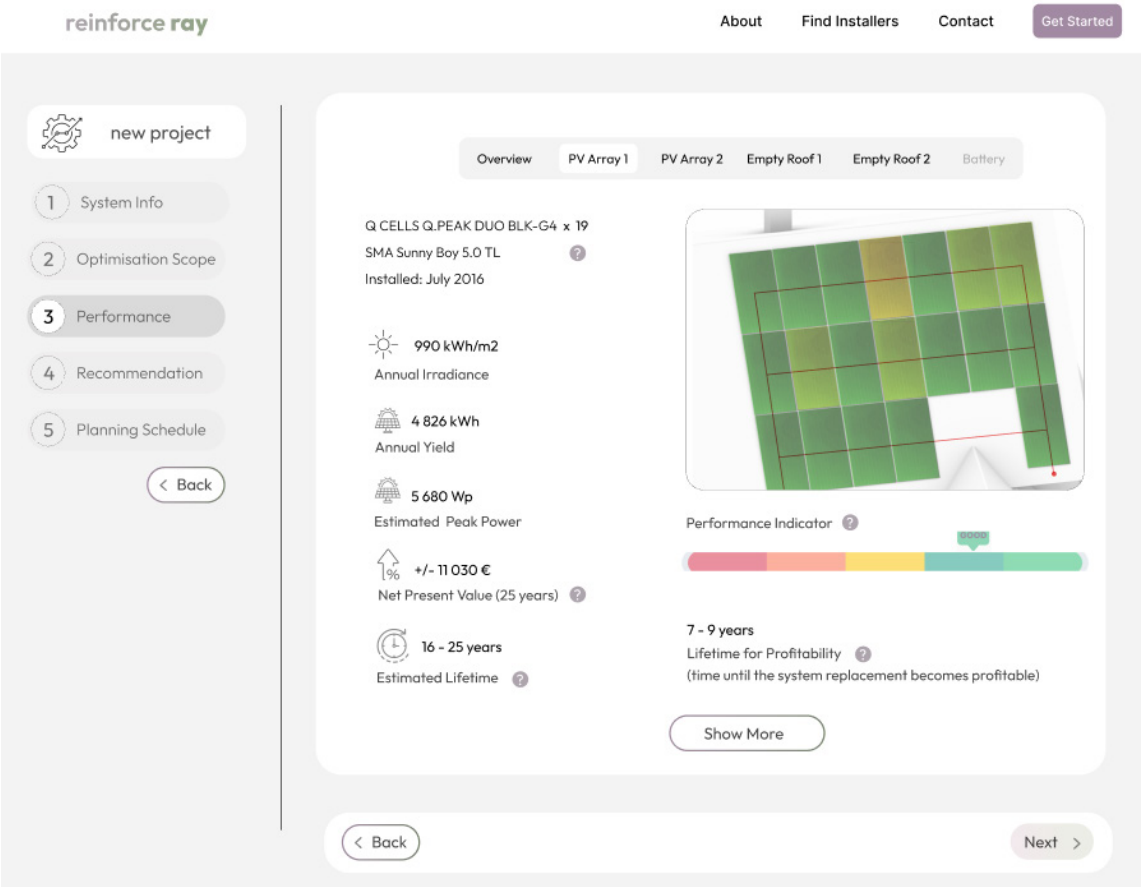


Figure 99: Performance overview - single array

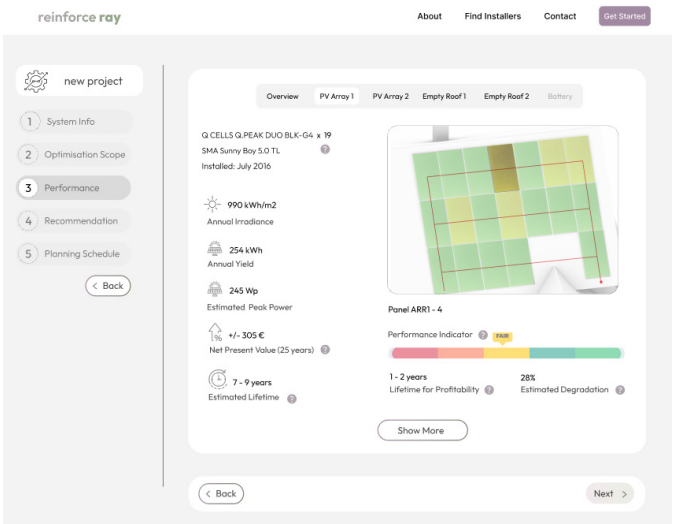


Figure 100: Performance overview - single component

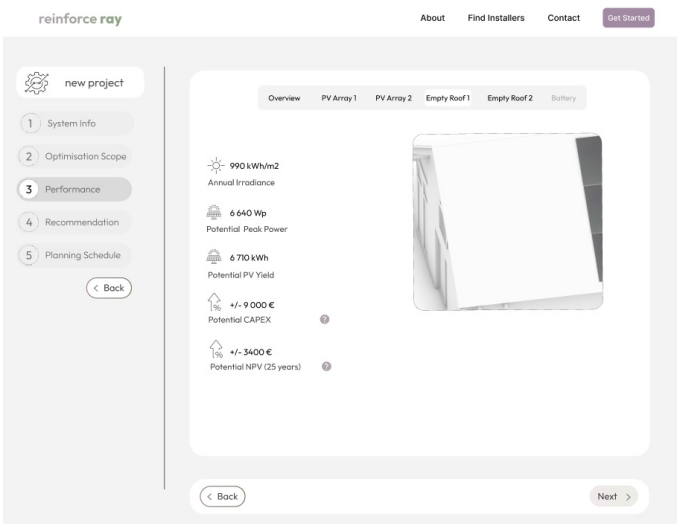


Figure 101: Performance overview - potential of a roof plane

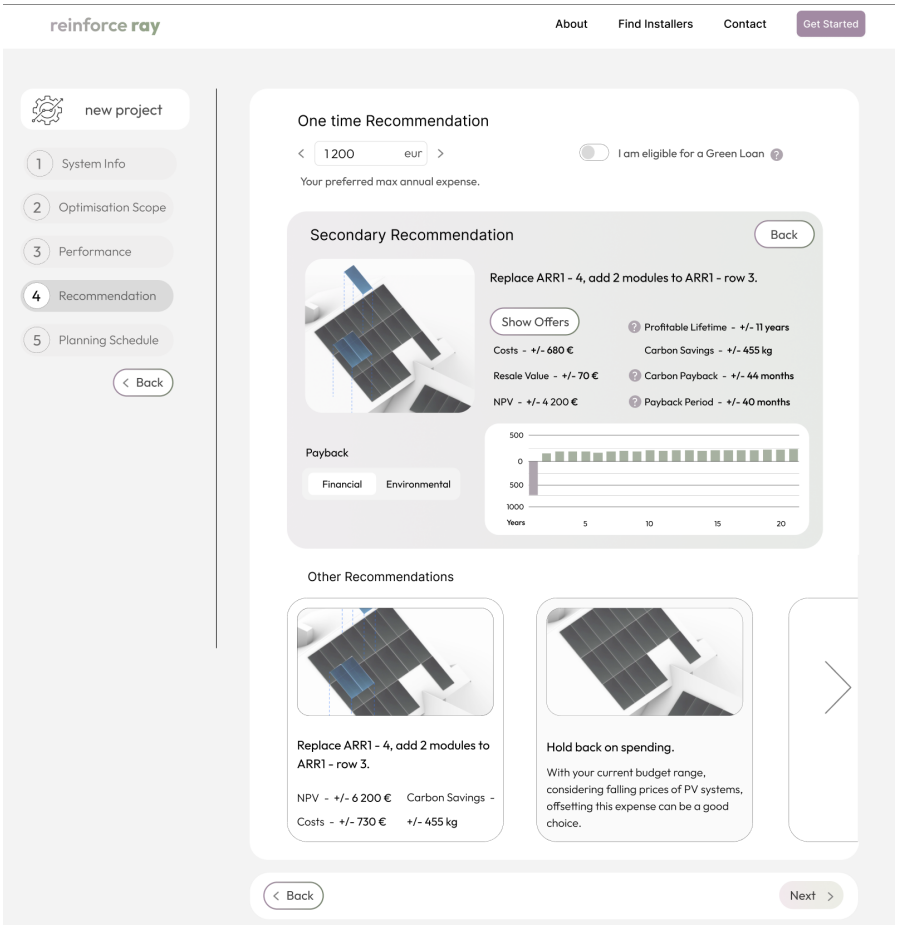


Figure 102: One-off secondary recommendation for a budget-constraint, non loan eligible and PV only scope - 1

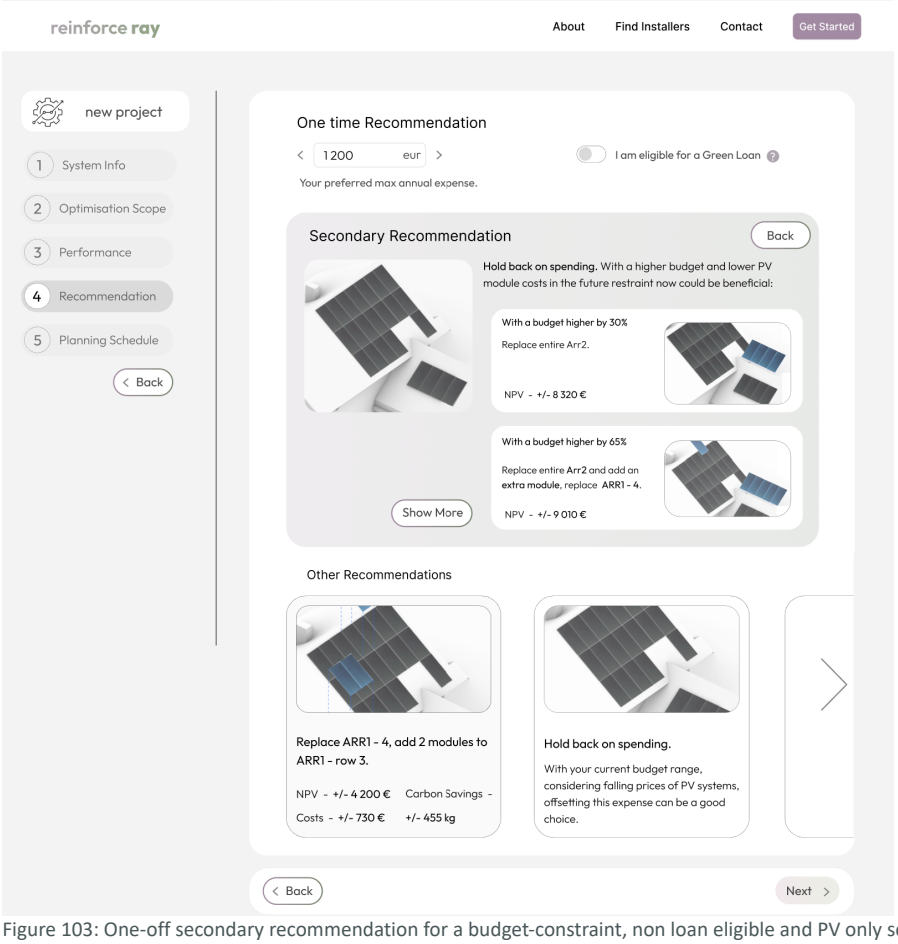


Figure 103: One-off secondary recommendation for a budget-constraint, non loan eligible and PV only scope - 2

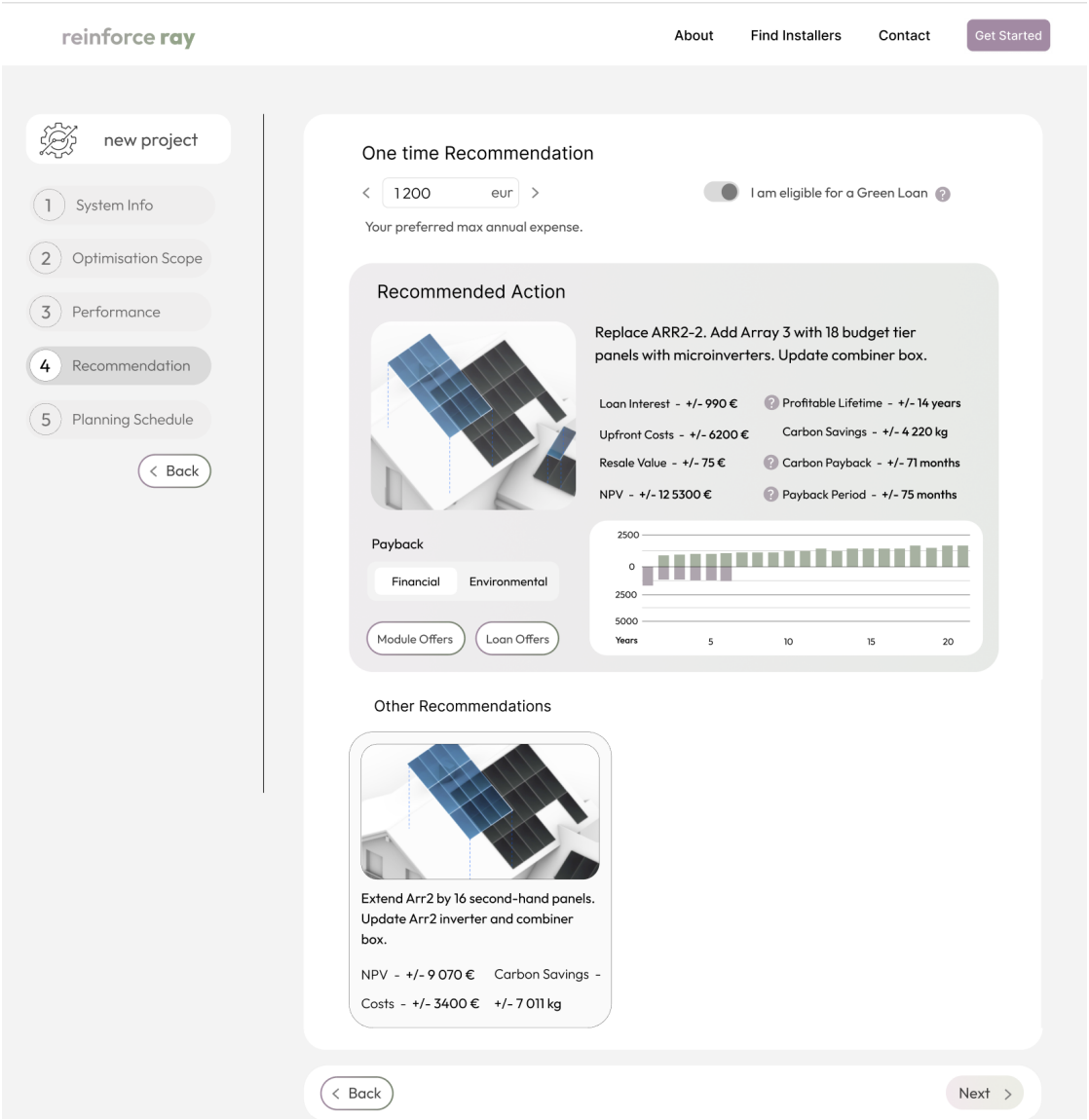


Figure 104: One-off recommendation for a budget-constraint, loan eligible and PV only scope

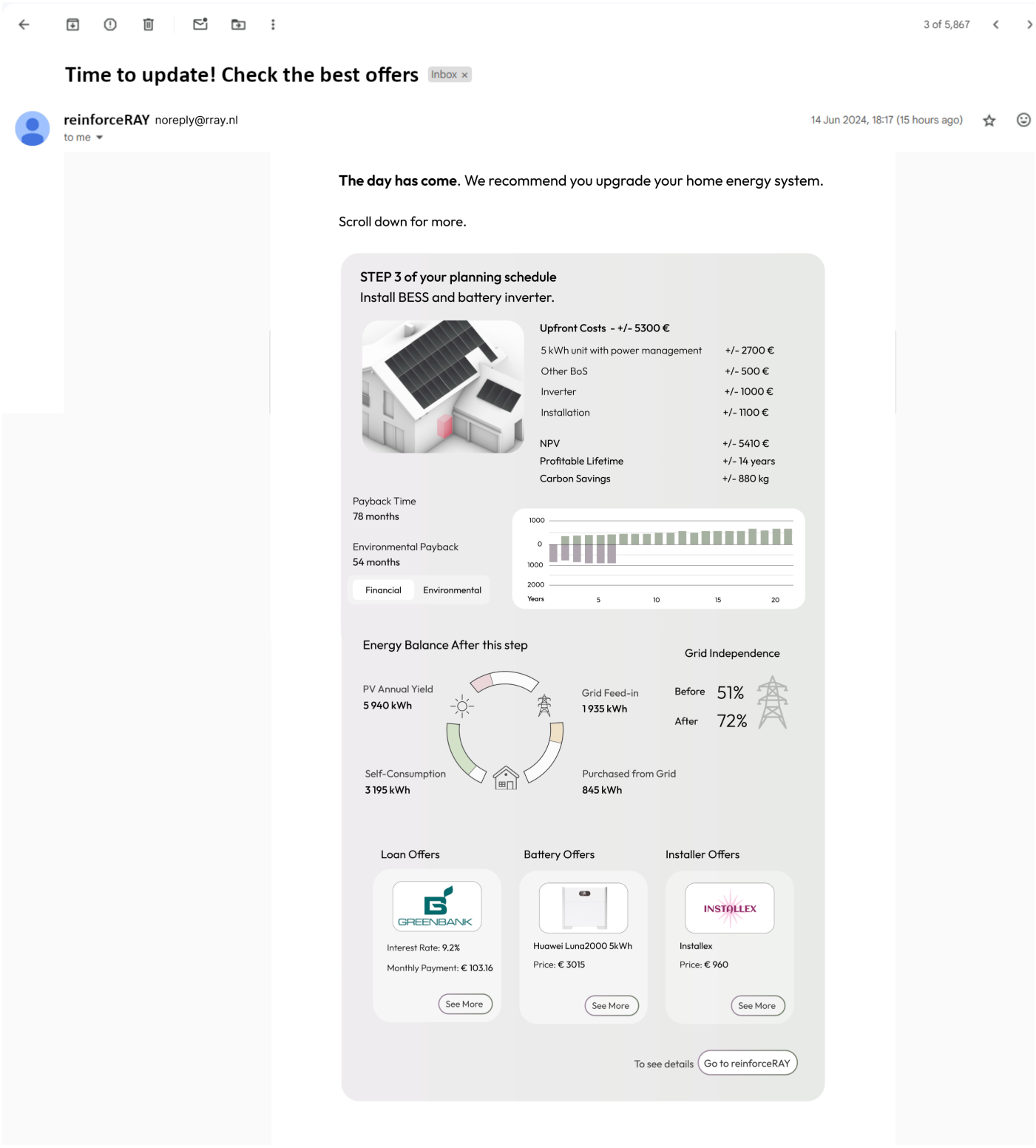


Figure 105: E-mail notification for the planning schedule step execution

REFERENCES

Abbid, A., Abdalla, A., Abid, A., Khan, D., Abdulrahman , A. and Zou, J. (2019). Gradio: Hassle-Free Sharing and Testing of ML Models in the Wild.

Abo-Khalil, A.G., Sayed, K., Radwan, A. and El-Sharkawy, I.I.A. (2023). Analysis of the PV System Sizing and Economic Feasibility Study in a grid-connected PV System. Case Studies in Thermal Engineering, 45. doi:https://doi.org/10.1016/j.csite.2023.102903.

Afridi, M., Tatapudi, S., Flicker, J., Srinivasan, D. and Govindasamy, T. (2023a). Reliability of Microinverters for Photovoltaic Systems: High-Temperature Accelerated Testing with Fixed and Cyclic Power Stresses. Energies, 16(18), pp.6511–6511. doi:https://doi.org/10.3390/en16186511.

Afridi, M., Tatapudi, S., Flicker, J., Srinivasan, D. and Tamizhmani, G. (2023b). Reliability of Microinverters for Photovoltaic Systems: High-Temperature Accelerated Testing with Fixed and Cyclic Power Stresses. Energies, 16(18), pp.6511–6511. doi:https://doi.org/10.3390/en16186511.

Afsar, M.M., Crump, T. and Far, B. (2022b). Reinforcement Learning based Recommender Systems: A Survey. ACM Computing Surveys. doi:https://doi.org/10.1145/3543846.

Akın Taşcıkaraoğlu and Ozan Erdinç (2019). Pathways to a smarter power system. London: Academic Press, pp.49–62.
Alexandros Iosifidis and Anastasios Tefas (2022). Deep Learning forRobot Perception and Cognition. Academic Press, pp.121–123.

Al Garni, H.Z., Awasthi, A. and Ramli, M.A.M. (2018). Optimal design and analysis of grid-connected photovoltaic under different tracking systems using HOMER. Energy Conversion and Management, 155, pp.42–57. doi:https://doi.org/10.1016/j.enconman.2017.10.090.

Aleksiejuk-Gawron, J., Milčiuvienė, S., Kiršienė, J., Doheijo, E., Garzon, D., Urbonas, R. and Milčius, D. (2020a). Net-Metering Compared to Battery-Based Electricity Storage in a Single-Case PV Application Study Considering the Lithuanian Context. Energies, 13(9), p.2286. doi:https://doi.org/10.3390/en13092286.

Alexandros Iosifidis and Anastasios Tefas (2022). Deep Learning for Robot Perception and Cognition. Academic Press, pp.121–123.

Allouhi, A. (2020). Solar PV integration in commercial buildings for self-consumption based on life-cycle economic/environmental multi-objective optimization. Journal of Cleaner Production, 270, p.122375. doi:https://doi.org/10.1016/j.jclepro.2020.122375.

An, Y., Xia, T., You, R., Lai, D., Liu, J. and Chen, C. (2021). A reinforcement learning approach for control of window behavior to reduce indoor PM2.5 concentrations in naturally ventilated buildings. Building and Environment, 200, pp.107978–107978. doi:https://doi.org/10.1016/j.buildenv.2021.107978.

Andreolli, F., D'Alpaos, C. and Moretto, M. (2021). Valuing Investments in Domestic PV-Battery Systems under Uncertainty. Energy Economics, 106, p.105721. doi:https://doi.org/10.1016/j.eneco.2021.105721.

Angenendt, G., Zurmühlen, S., Axelsen, H. and Sauer, D.U. (2018). Comparison of Different Operation Strategies for PV Battery Home Storage Systems Including forecast-based Operation Strategies. Applied Energy, 229, pp.884–899. doi:https://doi.org/10.1016/j.apenergy.2018.08.058.

Anzalchi, A. and Sarwat, A. (2017). Overview of technical specifications for grid-connected photovoltaic systems. Energy Conversion and Management, 152, pp.312–327. doi:https://doi.org/10.1016/j.enconman.2017.09.049.

Arar , A., Haouam , H., Rachid, C. and Aoun, N. (2019). Simplified Methods for Evaluating the Degradation of Photovoltaic Module and Modeling considering Partial Shading. Measurement, 138, pp.217–224. doi:https://doi.org/10.1016/j.measurement.2019.01.098.

Asango, L. (2018). Calibrating the Ornstein Uhlenbeck Equation Using Three Methods. African Institute for Mathematical Sciences, [Preprint].

Asif Hanif, M., Nadeem, F., Tariq, R. and Rashid, U. (2021). Renewable and Alternative Energy Resources. Academic Press.

Autarco (n.d.). Explore our solar monitoring solutions | Reliable and user-friendly | Autarco | Autarco. [online] www.autarco.com. Available at: https://www.autarco.com/en/products/monitoring [Accessed 4 Feb. 2024].

Barba, D. (2022). Stochastic Processes Simulation — the Ornstein Uhlenbeck Process. [online] Medium. Available at: https://

towardsdatascience.com/stochastic-processes-simulation-the-ornstein-uhlenbeck-process-e8bff820f3 [Accessed 26 Jan. 2024].

Beltran, H., Ayuso, P. and Pérez, E. (2020). Lifetime Expectancy of Li-Ion Batteries Used for Residential Solar Storage. Energies, 13(3), p.568. doi:https://doi.org/10.3390/en13030568.

Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J. and Zaremba, W. (2016). OpenAI Gym. doi:https://doi.org/10.48550/arXiv.1606.01540.

Brunton, S.L. and Jose Nathan Kutz (2022). Data-driven Science and Engineering: Machine learning, Dynamical systems, and Control. Cambridge, United Kingdom, New York, Ny: Cambridge University Press.

Butenko, A. (2016). Sharing Energy: Dealing with Regulatory Disconnect in Dutch Energy Law. SSRN Electronic Journal. doi:https://doi.org/10.2139/ssrn.2847590.

Carpinelli, G., Khormali, S., Mottola, F. and Proto, D. (2014). Battery Energy Storage Sizing When Time of Use Pricing Is Applied. The Scientific World Journal, 2014, pp.1–8. doi:https://doi.org/10.1155/2014/906284.

Centraal Bureau voor de Statistiek (2023). Energy consumption private dwellings; type of dwelling and regions. [online] Statistics Netherlands. Available at: https://www.cbs.nl/en-gb/figures/detail/81528ENG [Accessed 14 Jan. 2024].

Cervantes, J. and Choobineh, F. (2018). Optimal sizing of a nonutility-scale solar power system and its battery storage. Applied Energy, 216, pp.105–115. doi:https://doi.org/10.1016/j.apenergy.2018.02.013.

Chatzigeorgiou, N.G., Theocharides, S., Makrides, G. and Georghiou, G.E. (2024). A Review on Battery Energy Storage systems: Applications, developments, and Research Trends of Hybrid Installations in the end-user Sector. Journal of Energy Storage, 86(1).

Chello, A. (2020). A Gentle Introduction to Geometric Brownian Motion in Finance. [online] The Quant Journey. Available at: https://medium.com/the-quant-journey/a-gentle-introduction-to-geometric-brownian-motion-in-finance-68c37ba6f828.

Chen, X., Yao, L., McAuley, J., Zhou, G. and Wang, X. (2023a). Deep reinforcement learning in recommender systems: A survey and new perspectives. Knowledge-Based Systems, 264, p.110335. doi:https://doi.org/10.1016/j.knosys.2023.110335.

Chen, Z., Xiao, F., Guo, F. and Yan, J. (2023b). Interpretable Machine Learning for Building Energy management: a state-of-the-art Review. Advances in Applied Energy, [online] 9, p.100123. doi:https://doi.org/10.1016/j.adapen.2023.100123.

Cheng, Z., Zhao, Q., Wang, F., Jiang, Y., Xia, L. and Ding, J. (2016). Satisfaction Based Q-learning for Integrated Lighting and Blind Control. Energy and Buildings, 127, pp.43–55. doi:https://doi.org/10.1016/j.enbuild.2016.05.067.
Connolly, D., Lund, H., Mathiesen, B.V. and Leahy, M. (2010). A Review of Computer Tools for Analysing the Integration of Renewable Energy into Various Energy Systems. Applied Energy, 87(4), pp.1059–1082. doi:https://doi.org/10.1016/j.apenergy.2009.09.026.

Corte Real, A., Pontes Luz, G., Sousa, J.M.C., Brito, M.C. and Vieira, S.M. (2024). Optimization of a photovoltaic-battery System Using Deep Reinforcement Learning and Load Forecasting. Energy and AI, 16.

Deng, X., Wang, F., Hu, B., Lin, X. and Hu, X. (2022). Optimal Sizing of Residential Battery Energy Storage Systems for long-term Operational Planning. Journal of Power Sources, 551, p.232218. doi:https://doi.org/10.1016/j.jpowsour.2022.232218.

Dhimish, M. and Alrashidi, A. (2020). Photovoltaic Degradation Rate Affected by Different Weather Conditions: a Case Study Based on PV Systems in the UK and Australia. Electronics, 9(4), p.650. doi:https://doi.org/10.3390/electronics9040650.

Dhinakaran, A. (2021). A Look into Global, Cohort and Local Model Explainability. [online] Medium. Available at: https://towardsdatascience.com/a-look-into-global-cohort-and-local-model-explainability-973bd449969f.

E3P (2016). Typical Meteorological Year (TMY). [online] e3p.jrc.ec.europa.eu. Available at: https://e3p.jrc.ec.europa.eu/articles/typical-meteorological-year-tmy.

Eid, C., Reneses Guillén, J., Frías Marín, P. and Hakvoort, R. (2014). The economic effect of electricity net-metering with solar PV: Consequences for network cost recovery, cross subsidies and policy objectives. Energy Policy, 75, pp.244–254. doi:https://doi.org/10.1016/j.enpol.2014.09.011.

Ekici, S. and Kopru, M.A. (2016). Investigation of PV System Cable Losses. International Journal of Renewable Energy Research, 7(2).

Elshurafa, A.M., Albardi, S.R., Bigerna, S. and Bollino, C.A. (2018). Estimating the Learning Curve of Solar PV Balance–of–system for

over 20 countries: Implications and Policy Recommendations. Journal of Cleaner Production, [online] 196, pp.122–134. doi:https://doi.org/10.1016/j.jclepro.2018.06.016.

Enerdata (2024). 2050 Projections for CO2 Intensity of Electricity Generation. [online] eneroutlook.enerdata.net. Available at: https://eneroutlook.enerdata.net/forecast-world-co2-intensity-of-electricity-generation.html [Accessed 17 Mar. 2024].

Enphase (2024). Betrouwbare IQ 7+ omvormers voor zonne-energie | Enphase. [online] enphase.com. Available at: https://enphase.com/nl-nl/store/microinverters/iq7-series/iq7-microinverter [Accessed 26 Feb. 2024].

Erdinc, O., Paterakis, N.G., Pappi, I.N., Bakirtzis, A.G. and Catalão, J.P.S. (2015). A new perspective for sizing of distributed generation and energy storage for smart households under demand response. Applied Energy, 143, pp.26–37. doi:https://doi.org/10.1016/j.apenergy.2015.01.025.

Europe-Solarstore (2024). SunPower SPR-MAX2-360 Solar Panel. [online] www.europe-solarstore.com. Available at: https://www.europe-solarstore.com/solar-panels/manufacturer/sunpower/sunpower-spr-max2-360.html [Accessed 26 Feb. 2024].

European Commission (2023). Renewable Energy Targets. [online] energy.ec.europa.eu. Available at: https://energy.ec.europa.eu/topics/renewable-energy/renewable-energy-directive-targets-and-rules/renewable-energy-targets_en.

Eurostat (2023). Electricity and Gas Prices Stabilise in 2023. [online] ec.europa.eu. Available at: https://ec.europa.eu/eurostat/web/products-eurostat-news/w/ddn-20231026-1.
Eurostat (2024). Electricity price statistics - Statistics Explained. [online] ec.europa.eu. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Electricity_price_statistics.

Faiman, D. (2008). Assessing the outdoor operating temperature of photovoltaic modules. Progress in Photovoltaics: Research and Applications, 16(4), pp.307–315. doi:https://doi.org/10.1002/pip.813.

Fan, J., Zheng, Q. and Wang, J. (2018). Photovoltaic Modules Power Degradation and Lifetime Prediction under Accelerated Damp-heat Conditions Based on Gamma Process. International Conference on Power System Technology (POWERCON). doi:https://doi.org/10.1109/powercon.2018.8601560.

Fan, Y. and Xia, X. (2017). A multi-objective optimization model for energy-efficiency building envelope retrofitting plan with rooftop PV system installation and maintenance. Applied Energy, 189, pp.327–335. doi:https://doi.org/10.1016/j.apenergy.2016.12.077.

Farida Agustini, W., Affianti, I.R. and Putri, E.R. (2018). Stock Price Prediction Using Geometric Brownian Motion. Journal of Physics: Conference Series, 974, p.012047. doi:https://doi.org/10.1088/1742-6596/974/1/012047.

Fraunhofer Institute for Solar Energy Systems (2023). Photovoltaics Report. Freiburg.

Frischknecht, R., Itten, R. and Wyss, F. (2015). Life Cycle Assessment of Future Photovoltaic Electricity Production from Residential-scale Systems Operated in Europe. IEA-PVPS.

Fu, Q., Han, Z., Chen, J., Lu, Y., Wu, H. and Wang, Y. (2022). Applications of Reinforcement Learning for Building Energy Efficiency control: a Review. Journal of Building Engineering, 50, p.104165. doi:https://doi.org/10.1016/j.jobe.2022.104165.

Galatro, D., Romero, D.A., Freitez, J., Da Silva, C., Trescases, O. and Amón, C.H. (2021). Modeling Degradation of lithium-ion Batteries considering cell-to-cell Variations. Journal of Energy Storage, 44, pp.103478–103478. doi:https://doi.org/10.1016/j.est.2021.103478.

Gao, G., Li, J. and Wen, Y. (2020). DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings via Reinforcement Learning. IEEE Internet of Things Journal, 7(9), pp.8472–8484. doi:https://doi.org/10.1109/jiot.2020.2992117.

Gerbinet, S., Belboom, S. and Léonard, A. (2014). Life Cycle Analysis (LCA) of Photovoltaic panels: a Review. Renewable and Sustainable Energy Reviews, 38, pp.747–753. doi:https://doi.org/10.1016/j.rser.2014.07.043.

Golombek, R., Lind, A., Ringkjøb, H.-K. and Seljom, P. (2022). The Role of Transmission and Energy Storage in European Decarbonization Towards 2050. Energy, 239, p.122159. doi:https://doi.org/10.1016/j.energy.2021.122159.
Google PAIR. (2019). People + AI Guidebook. 2nd ed. People + AI Research.

Guan, C., Wang, Y., Lin, X., Nazarian, S. and Pedram, M. (2015). Reinforcement Learning-Based Control of Residential Energy Storage

Systems for Electric Bill Minimization. In: Consumer Communications and Networking Conference, CCNC IEEE. 2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC).

Guerra, G., Pau Mercade-Ruiz, Gaetana Anamiati and Landberg, L. (2023). Long-term PV System Modelling and Degradation Using Neural Networks. EPJ Photovoltaics, 14, pp.30–30. doi:https://doi.org/10.1051/epjpv/2023018.

Gyamfi, S., Aboagye, B., Peprah, F. and Obeng, M. (2023). Degradation Analysis of Polycrystalline Silicon Modules from Different Manufacturers under the Same Climatic Conditions. Energy conversion and management. X, 20, pp.100403–100403. doi:https://doi.org/10.1016/j.ecmx.2023.100403.

Han, A. (2014). Efficiency of Solar PV, then, Now and Future – Solar Photovoltaic. [online] Lafayette.edu. Available at: https://sites.lafayette.edu/egrs352-sp14-pv/technology/history-of-pv-technology/.

Hartner, M., Mayr, D., Kollmann, A. and Haas, R. (2017). Optimal sizing of residential PV-systems from a household and social cost perspective. Solar Energy, 141, pp.49–58. doi:https://doi.org/10.1016/j.solener.2016.11.022.

Hashemi, B., Taheri, S., Cretu, A.-M. and Pouresmaeil, E. (2021). Systematic Photovoltaic System Power Losses Calculation and Modeling Using Computational Intelligence Techniques. Applied Energy, 284, p.116396. doi:https://doi.org/10.1016/j.apenergy.2020.116396.

Hayat, M.A., Shahnia, F. and Shafiullah, G. (2019). Replacing Flat Rate Feed-In Tariffs for Rooftop Photovoltaic Systems with a Dynamic One to Consider Technical, Environmental, Social, and Geographical Factors. IEEE Transactions on Industrial Informatics, 15(7), pp.3831–3844. doi:https://doi.org/10.1109/tii.2018.2887281.

Hayes, C.F., Rădulescu, R., Bargiacchi, E., Källström, J., Macfarlane, M., Reymond, M., Verstraeten, T., Zintgraf, L.M., Dazeley, R., Heintz, F., Howley, E., Irissappane, A.A., Mannion, P., Nowé, A., Ramos, G., Restelli, M., Vamplew, P. and Roijers, D.M. (2022). A practical guide to multi-objective reinforcement learning and planning. Autonomous Agents and Multi-Agent Systems, 36(1). doi:https://doi.org/10.1007/s10458-022-09552-y.

Hazim Imad Hazim, Kyairul Azmi Baharin, Chin Kim Gan, Sabry, A.H. and Humaidi, A.J. (2023). Review on Optimization Techniques of PV/Inverter Ratio for Grid-Tie PV Systems. Applied Sciences, 13(5), pp.3155–3155. doi:https://doi.org/10.3390/app13053155.

He, L., Li, C.-L., Nie, Q.-Y., Men, Y., Shao, H. and Zhu, J. (2017). Core Abilities Evaluation Index System Exploration and Empirical Study on Distributed PV-Generation Projects. Energies, 10(12), p.2083. doi:https://doi.org/10.3390/en10122083.

Holmgren, W.F. et al. (2015) 'PVLIBPython 2015', 2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)[Preprint]. doi:10.1109/pvsc.2015.7356005.

Hosseini, S., Taheri, S., Farzaneh, M., Taheri, H. and Narimani, M. (2018). Determination of Photovoltaic Characteristics in Real Field Conditions. IEEE Journal of Photovoltaics, 8(2), pp.572–580. doi:https://doi.org/10.1109/jphotov.2018.2797974.

Ibe, O.C. (2013). Markov Processes for Stochastic Modeling. San Diego, CA: Elsevier Science, pp.329–347.

Ioannis Tsiropoulos, Wouter Nijs, Dalius Tarvydas and Ruiz Castello Pablo (2020). Towards net-zero Emissions in the EU Energy System by 2050. Publications Office of the European Union. doi:https://doi.org/10.2760/081488.

IRENA (2022). Renewable Power Generation Costs in 2022. International Renewable Energy Agency.

IRENA (2023). Renewable Power Generation Costs in 2022. Abu Dhabi: International Renewable Energy Agency.

Jäger-Waldau , A. (2016). Costs and Economics of Electricity from Residential PV Systems in Europe. [online] www.europeanenergyinnovation.eu. Available at: https://www.europeanenergyinnovation.eu/Articles/Winter-2016/Costs-and-Economics-of-Electricity-from-Residential-PV-Systems-in-Europe [Accessed 9 Apr. 2024].

Javeed, I., Khezri, R., Mahmoudi, A., Yazdani, A. and Shafiullah, G.M. (2021). Optimal Sizing of Rooftop PV and Battery Storage for Grid-Connected Houses Considering Flat and Time-of-Use Electricity Rates. Energies, 14(12), p.3520. doi:https://doi.org/10.3390/en14123520.

Jo, E.-H. and Lee, H.-J. (2019). Variation of Solar Photovoltaic Power Estimation Due to Solar Irradiance Decomposition Models. Journal of the Korean Solar Energy Society, 39(3), pp.81–99. doi:https://doi.org/10.7836/kses.2019.39.3.081.

REFERENCES

Johnen, M., Pitzen, S., Kamps, U., Kateri, M., Dechent, P. and Sauer, D.U. (2021). Modeling long-term Capacity Degradation of lithium-ion Batteries. Journal of Energy Storage, 34, p.102011. doi:https://doi.org/10.1016/j.est.2020.102011.

Joint Research Centre (2022). Photovoltaics in the European Union. Ispra, Italy: European Commission.

Jung, S., Jeoung, J., Kang, H. and Hong, T. (2021). Optimal Planning of a Rooftop PV System Using GIS-based Reinforcement Learning. Applied Energy, 298(117239). doi:https://doi.org/10.1016/j.apenergy.2021.117239.

Jung, S., Kang, H., Lee, M. and Hong, T. (2020). An optimal scheduling model of an energy storage system with a photovoltaic system in residential buildings considering the economic and environmental aspects. Energy and Buildings, 209, p.109701. doi:https://doi.org/10.1016/j.enbuild.2019.109701.

Kambezidis, H.D. and Psiloglou, B.E. (2021). Estimation of the Optimum Energy Received by Solar Energy Flat-Plate Convertors in Greece Using Typical Meteorological Years. Part I: South-Oriented Tilt Angles. Applied Sciences, 11(4), p.1547. doi:https://doi.org/10.3390/app11041547.

Khanal, S., Khezri, R., Mahmoudi, A. and Kahourzadeh, S. (2023). Optimal Capacity of Solar Photovoltaic and Battery Storage for Grid-tied Houses Based on Energy Sharing. Iet Generation Transmission & Distribution, 17(8), pp.1707–1722. doi:https://doi.org/10.1049/gtd2.12824.

Khanfara, M. (2018). Economic Sizing of a Grid-Connected Photovoltaic System: Case of GISER Research Project in Morocco. International Conference on Renewable Energies and Energy Efficiency, 161(10.1088/1755-1315/161/1/012006).

Khezri, R., Mahmoudi, A. and Aki, H. (2022). Optimal Planning of Solar Photovoltaic and Battery Storage Systems for grid-connected Residential sector: Review, Challenges and New Perspectives. Renewable and Sustainable Energy Reviews, 153, p.111763. doi:https://doi.org/10.1016/j.rser.2021.111763.

Khezri, R., Mahmoudi, A. and Haque, M.H. (2020). Optimal Capacity of Solar PV and Battery Storage for Australian Grid-Connected Households. IEEE Transactions on Industry Applications, 56(5), pp.5319–5329. doi:https://doi.org/10.1109/tia.2020.2998668.

Khoury, J., Mbayed, R., Salloum, G. and Monmasson, E. (2015). Optimal Sizing of a Residential PV-battery Backup for an Intermittent Primary Energy Source under Realistic Constraints. Energy and Buildings, 105, pp.206–216. doi:https://doi.org/10.1016/j.enbuild.2015.07.045.

Kim, H., Cheon, H., Ahn, Y.-H. and Choi, D.G. (2019). Uncertainty quantification and scenario generation of future solar photovoltaic price for use in energy system models. Energy, 168, pp.370–379. doi:https://doi.org/10.1016/j.energy.2018.11.075.

Kimber, A., Mitchell, L., Nogradi, S. and Wenger, H. (2006). The Effect of Soiling on Large Grid-Connected Photovoltaic Systems in California and the Southwest Region of the United States. [online] IEEE Xplore. doi:https://doi.org/10.1109/WCPEC.2006.279690.

King, D.L., Gonzalez, S., Galbraith, G.M. and Boyson, W.E. (2007). Performance Model for Grid-connected Photovoltaic Inverters. Albuquerque, New Mexico: Sandia National Laboratories.

Klaassen, E.A.M., Frunt, J. and Slootweg, H. (2015). Assessing the Impact of Distributed Energy Resources on LV Grids Using Practical Measurements.

Klushin, P. (2024). What does it take to deploy ML models in production? [online] www.qwak.com. Available at: https://www.qwak.com/post/what-does-it-take-to-deploy-ml-models-in-production.

Kornelakis, A. and Marinakis, Y. (2010). Contribution for optimal sizing of grid-connected PV-systems using PSO. Renewable Energy, 35(6), pp.1333–1341. doi:https://doi.org/10.1016/j.renene.2009.10.014.

Kouro, S., Leon, J.I., Vinnikov, D. and Franquelo, L.G. (2015). Grid-Connected Photovoltaic Systems: An Overview of Recent Research and Emerging PV Converter Technology. IEEE Industrial Electronics Magazine, 9(1), pp.47–61. doi:https://doi.org/10.1109/mie.2014.2376976.

Krebs, L., Frischknecht, R. and Stolz, P. (2020). Environmental Life Cycle Assessment of Residential PV and Battery Storage Systems. International Energy Agency.

Kucuksari, S., Khaleghi, A.M., Hamidi, M., Zhang, Y., Szidarovszky, F., Bayraksan, G. and Son, Y.-J. (2014). An Integrated GIS, Optimization and Simulation Framework for Optimal PV Size and Location in Campus Area Environments. Applied Energy, 113,

pp.1601–1613. doi:https://doi.org/10.1016/j.apenergy.2013.09.002.

Kuitche,J. M. (2010). 'Statistical Lifetime Prediction for Photovoltaic Modules'[PowerPoint presentation]. Arizona State University& TUV Rheinland PTL. Available at: https://www.energy.gov/eere/office-energy-efficiency-renewable-energy(Accessed: 20 January 2024).

Kumar, S. and Sarkar, B. (2013). Design for Reliability with Weibull Analysis for Photovoltaic Modules. International Journal of Current Engineering and Technology, 3(1).

Kurdi, Y., Alkhatatbeh, B.J., Asadi, S. and Jebelli, H. (2022). A decision-making Design Framework for the Integration of PV Systems in the Urban Energy Planning Process. Renewable Energy, 197, pp.288–304. doi:https://doi.org/10.1016/j.renene.2022.07.001.

Lai, G., Wang, D., Wang, Z., Fan, F., Wang, Q. and Wang, R. (2023a). Distribution-based PV Module Degradation Model. Energy Science & Engineering, 11(3). doi:https://doi.org/10.1002/ese3.1401.

Lau, K.Y., Muhamad, N.A., Arief, Y.Z., Tan, C.W. and Yatim, A.H.M. (2016). Grid-connected photovoltaic systems for Malaysian residential sector: Effects of component costs, feed-in tariffs, and carbon taxes. Energy, 102, pp.65–82. doi:https://doi.org/10.1016/j.energy.2016.02.064.

Lee, M. and Hong, T. (2019). Hybrid agent-based modeling of rooftop solar photovoltaic adoption by integrating the geographic information system and data Mining technique. Energy Conversion and Management, 183, pp.266–279. doi:https://doi.org/10.1016/j.enconman.2018.12.096.

Leone, U. (2023). Energy Company Vandebron Charge Customers with Solar Panels around €20 Extra. [online] nltimes.nl. Available at: https://nltimes.nl/2023/08/15/energy-company-vandebron-charge-customers-solar-panels-around-eu20-extra [Accessed 14 Jan. 2024].

Liu, C., Jing Yang, R., Yu, X., Sun, C., Rosengarten, G., Liebman, A., Wakefield, R., Wong, P.S. and Wang, K. (2023). Supporting Virtual Power Plants decision-making in Complex Urban Environments Using Reinforcement Learning. Sustainable Cities and Society, 99, pp.104915–104915. doi:https://doi.org/10.1016/j.scs.2023.104915.

Liu, G., Rasul, M.G., Amanullah, M.T.O. and Khan, M.M.K. (2012). Techno-economic simulation and optimization of residential grid-connected PV system for the Queensland climate. Renewable Energy, 45, pp.146–155. doi:https://doi.org/10.1016/j.renene.2012.02.029.

Luo, F., Ranzi, G., Wang, S. and Dong, Z.Y. (2019). Hierarchical Energy Management System for Home Microgrids. IEEE Transactions on Smart Grid, 10(5), pp.5536–5546. doi:https://doi.org/10.1109/tsg.2018.2884323.

Luo, X., Jorge Varela Barreras, Chambon, C.L., Wu, B. and Efstratios Batzelis (2021). Hybridizing Lead–Acid Batteries with Supercapacitors: a Methodology. Energies, 14(2), pp.507–507. doi:https://doi.org/10.3390/en14020507.

Ma, M., Ma, J., Wang, H., Ma, W. and Zhang, R. (2022). PV Module Life Prediction Based on Coupled Failure Model. Microelectronics Reliability, p.114739. doi:https://doi.org/10.1016/j.microrel.2022.114739.

Maisch, M. (2023). Europe's Big Battery Fleet to Surge to 95 GW by 2050, Says Research Firm. [online] Pv Magazine International. Available at: https://www.pv-magazine.com/2023/04/24/europes-big-battery-fleet-to-surge-to-95-gw-by-2050-says-research-firm/ [Accessed 4 May 2024].

Merton, R.C. (1976). Option Pricing When Underlying Stock Returns Are Discountinous. Journal of Financial Economics, 3, pp.125–144.

Ministerie van Economische Zaken en Klimaat (2023). Group of European Countries Aim to Decarbonize Their Electricity System by 2035 - News Item - Government.nl. [online] www.government.nl. Available at: https://www.government.nl/latest/news/2023/12/18/group-of-european-countries-aim-to-decarbonize-their-electricity-system-by-2035.

Mondol, J.D., Yohanis, Y.G. and Norton, B. (2009). Optimising the economic viability of grid-connected photovoltaic systems. Applied Energy, 86(7-8), pp.985–999. doi:https://doi.org/10.1016/j.apenergy.2008.10.001.

Moon, Y. and Baran, M. (2018). Economic analysis of a residential PV system from the timing perspective: A real option model.

REFERENCES

Renewable Energy, 125, pp.783–795. doi:https://doi.org/10.1016/j.renene.2018.02.138.

Morales, R., Marín, L.G., Roje, T., Caquilpan, V., Sáez, D. and Nuñez, A. (2024). Microgrid Planning Based on Computational Intelligence Methods for Rural communities: a Case Study in the José Painecura Mapuche community, Chile. Expert Systems with Applications, 235, pp.121179–121179. doi:https://doi.org/10.1016/j.eswa.2023.121179.

Morey, M., Gupta, N. and kumar, A. (2023). A comprehensive review of grid-connected solar photovoltaic system: Architecture, control, and ancillary services. Renewable Energy Focus, 45, pp.307–330. doi:https://doi.org/10.1016/j.ref.2023.04.009.

Nafeh, A.E.-S.A. (2009). Design and Economic Analysis of a Stand-Alone PV System to Electrify a Remote Area Household in Egypt. The Open Renewable Energy Journal, 2(1), pp.33–37. doi:https://doi.org/10.2174/1876387100902010033.

Narayan, N., Papakosta, T., Vega-Garita, V., Qin, Z., Popovic-Gerber, J., Bauer, P. and Zeman, M. (2018). Estimating Battery Lifetimes in Solar Home System Design Using a Practical Modelling Methodology. Applied Energy, 228, pp.1629–1639. doi:https://doi.org/10.1016/j.apenergy.2018.06.152.

National Renewable Energy Laboratory (2022). PV Fleet Performance Data Initiative. [online] www.nrel.gov. Available at: https://www.nrel.gov/pv/fleet-performance-data-initiative.html [Accessed 19 Mar. 2024].

Nehme, B., M'Sirdi, N.K., Akiki, T. and Zeghondy, B. (2020). Assessing the Effect of Temperature on Degradation Modes of PV Panels. In: International Conference on Renewable Energies for Developing Countries .

Ngoc An, L., Quoc-Tuan, T., Seddik, B. and Van-Linh, N. (2015). Optimal Sizing of a grid-connected Microgrid. In: IEEE International Conference on Industrial Technology (ICIT). 2015 IEEE International Conference on Industrial Technology (ICIT).
Norouzi, F., Shekhar, A., Hoppe, T. and Bauer, P. (2023). Economic Impact of New Pricing Policies on Solar PV Households in the Netherlands. In: Innovative Smart Grid Technologies Europe (ISGT EUROPE).

Ogunjuyigbe, A.S.O., Ayodele, T.R. and Akinola, O.A. (2016). Optimal Allocation and Sizing of PV/Wind/Split-diesel/Battery Hybrid Energy System for Minimizing Life Cycle cost, Carbon Emission and Dump Energy of Remote Residential Building. Applied Energy, 171, pp.153–171. doi:https://doi.org/10.1016/j.apenergy.2016.03.051.

Onile, A.E., Belikov, J., Levron, Y. and Petlenkov, E. (2022). Energy Efficient Behavior Modelling for Demand Side Recommender System in Solar Microgrid Applications Using Multi Agent Reinforcement Learning Model. SSRN Electronic Journal, [Preprint]. doi:https://doi.org/10.2139/ssrn.4196519.

Ozcan, H.G., Gunerhan, H., Yildirim, N. and Hepbasli, A. (2019). A comprehensive evaluation of PV electricity production methods and life cycle energy-cost assessment of a particular system. Journal of Cleaner Production, 238, p.117883. doi:https://doi.org/10.1016/j.jclepro.2019.117883.

Park, J.Y., Dougherty, T., Fritz, H. and Nagy, Z. (2019). LightLearn: an Adaptive and Occupant Centered Controller for Lighting Based on Reinforcement Learning. Building and Environment, 147, pp.397–414. doi:https://doi.org/10.1016/j.buildenv.2018.10.028.

Paudyal, B.R. and Imenes, A.G. (2021). Investigation of Temperature Coefficients of PV Modules through Field Measured Data. Solar Energy, 224, pp.425–439. doi:https://doi.org/10.1016/j.solener.2021.06.013.

Penizzotto, F., Pringles, R. and Olsina, F. (2019). Real Options Valuation of Photovoltaic Power Investments in Existing Buildings. Renewable and Sustainable Energy Reviews, 114(109308). doi:https://doi.org/10.1016/j.rser.2019.109308.

Perez, R., Ineichen, P., Seals, R., Michalsky, J. and Stewart, R. (1990). Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy, 44(5), pp.271–289. doi:https://doi.org/10.1016/0038-092x(90)90055-h.

Perpetuo e Oliveira, T., Inga Narváez, D. and Gradella Villalva, M. (2019). Comparison of Irradiance Decomposition and Energy Production Methods in a Solar Photovoltaic System. Zenodo (CERN European Organization for Nuclear Research), 13(5). doi:https://doi.org/10.5281/zenodo.3298751.

Petters, A.O. and Dong , X. (2016). Stochastic Calculus and Geometric Brownian Motion Model. In: An Introduction to Mathematical Finance with Applications. New York: Springer.

Polanitzer, R. (2023). Geometric Brownian Motion (Random Walk) Process with Drift in Python; Simulate the Future…. [online] Medium. Available at: https://medium.com/@polanitzer/path-dependent-brownian-motion-random-walk-process-in-python-

simulate-the-future-distribution-of-3e13f1ef5e2d [Accessed 13 Feb. 2024].

Price, B. (2024). Second Hand Solar Panels - Are Used Panels Worth It? [online] heatable.co.uk. Available at: https://heatable.co.uk/solar/advice/second-hand-solar-panels [Accessed 21 Apr. 2024].

pvlib (2023) API reference, APIreference - pvlib python 0.10.3 documentation. Available at:https://pvlib-python.readthedocs.io/en/stable/reference/index.html (Accessed:13 January 2024).

Qi , B., Rashedi, M. and Ardakanian, O. (2019). EnergyBoost: Learning-based Control of Home Batteries. In: e-Energy '19: Proceedings of the Tenth ACM International Conference on Future Energy Systems. New York: Association for Computing Machinery.

Raffin, A., Hill, A., Gleave, A., Kanervisto, A., Ernestus, M. and Dormann, N. (2021). Stable-Baselines3: Reliable Reinforcement Learning Implementations. Journal of Machine Learning Research, 22, pp.1–8.

Rahman, T., Al Mansur, A., Hossain Lipu, M.S.H., Lipu, H., Ashique, R.H., Houran, M.A., Elavarasan, R.M. and Hossain, E. (2023). Investigation of Degradation of Solar Photovoltaics: A Review of Aging Factors, Impacts, and Future Directions toward Sustainable Energy Management. Energies, 16(9).

Ramli, M.A.M., Bouchekara, H.R.E.H. and Alghamdi, A.S. (2018). Optimal sizing of PV/wind/diesel hybrid microgrid system using multi-objective self-adaptive differential evolution algorithm. Renewable Energy, 121, pp.400–411. doi:https://doi.org/10.1016/j.renene.2018.01.058.

Renewables Now (2022). Europe Expected to Have 32.2 GWh of Home Batteries by end-2026. [online] 247 Energy Storage. Available at: https://247storage.energy/europe-expected-to-have-32-2-gwh-of-home-batteries-by-end-2026/ [Accessed 27 Apr. 2024].

Rijksdienst voor Ondernemend Nederland (2024). Elektriciteit Totaal incl. Btw [€/kWh] Nederland.

Samsukha, A. (2023). Best Solar Panel Monitoring Software. [online] Tech Blog | Mobile App, eCommerce, Salesforce Insights. Available at: https://www.emizentech.com/blog/best-solar-monitoring-system.html [Accessed 4 Feb. 2024].

Scarlat, N., Prussi, M. and Padella, M. (2022). Quantification of the carbon intensity of electricity produced and used in Europe. Applied Energy, 305, p.117901. doi:https://doi.org/10.1016/j.apenergy.2021.117901.

Scheepers, M., Palacios, S.G., Jegu, E., Nogueira, L.P., Rutten, L., van Stralen, J., Smekens, K., West, K. and van der Zwaan, B. (2022). Towards a climate-neutral Energy System in the Netherlands. Renewable and Sustainable Energy Reviews, [online] 158, p.112097. doi:https://doi.org/10.1016/j.rser.2022.112097.

Schlemminger, M., Ohrdes, T., Schneider, E. and Knoop, M. (2022). Dataset on Electrical single-family House and Heat Pump Load Profiles in Germany. Scientific Data, [online] 9(1), p.56. doi:https://doi.org/10.1038/s41597-022-01156-1.
Schulman, J., Wolski, F., Dharival, P., Radford, A. and Klimov, O. (2022). Proximal Policy Optimization Algorithms .

Schulte, E., Scheller, F., Sloot, D. and Bruckner, T. (2022). A meta-analysis of Residential PV adoption: the Important Role of Perceived benefits, Intentions and Antecedents in Solar Energy Acceptance. Energy Research & Social Science, 84, p.102339. doi:https://doi.org/10.1016/j.erss.2021.102339.
Seppo Sierla, Heikki Ihasalo and Valeriy Vyatkin (2022). A Review of Reinforcement Learning Applications to Control of Heating, Ventilation and Air Conditioning Systems. Energies, 15(10), pp.3526–3526. doi:https://doi.org/10.3390/en15103526.

Sharma, P., Kolhe, M. and Sharma, A. (2019). Economic Analysis of a Building Integrated Photovoltaic System Without and With Energy Storage. IOP Conference Series: Materials Science and Engineering, 605, p.012013. doi:https://doi.org/10.1088/1757-899x/605/1/012013.

Shiva Gorjian and Ashish Shukla (2020a). Photovoltaic solar energy conversion : technologies applications and environmental impacts. London: Academic Press.

Shiva Gorjian and Ashish Shukla (2020b). Photovoltaic Solar Energy Conversion: Technologies Applications and Environmental Impacts. London: Academic Press, pp.313–346.

Shongwe, S. and Hanif, M. (2015). Comparative Analysis of Different Single-Diode PV Modeling Methods. IEEE Journal of Photovoltaics, 5(3), pp.938–946. doi:https://doi.org/10.1109/jphotov.2015.2395137.

Shuvo, S.S. and Yilmaz, Y. (2022). Home Energy Recommendation System (HERS): a Deep Reinforcement Learning Method Based on Residents' Feedback and Activity. IEEE Transactions on Smart Grid, 13(4), pp.1–1. doi:https://doi.org/10.1109/tsg.2022.3158814.

Siewierski, T. and Malarek, A. (2022). Optimization of Battery Storage Capacity and Operation for Balancing Residential PV Installation. In: International Conference on Software, Knowledge Information, Industrial Management and Applications (SKIMA).

Simpkins, T., Anderson, K., Cutler, D. and Olis, D. (2016). Optimal Sizing of a solar-plus-storage System for Utility Bill Savings and Resiliency Benefits. In: OSTI OAI (U.S. Department of Energy Office of Scientific and Technical Information). Office of Scientific and Technical Information. doi:https://doi.org/10.1109/isgt.2016.7781237.

Smith, Z.A. and Taylor, K.D. (2008). Renewable and Alternative Energy Resources: a Reference Handbook. Santa Barbara, California: Abc-Clio, pp.241–243.

SolarPower Europe (2023a). EU Market Outlook for Solar Power 2023-2027. [online] www.solarpowereurope.org. Available at: https://www.solarpowereurope.org/insights/outlooks/eu-market-outlook-for-solar-power-2023-2027/detail.

SolarPower Europe (2023b). Top 10 EU Countries Solar Capacity per Capita. [online] www.solarpowereurope.org. Available at: https://www.solarpowereurope.org/advocacy/solar-saves/fact-figures/top-10-eu-countries-solar-capacity.

Stander, J. (2020). Brownian Motion Share Price Modelling. [online] Youtube.com. Available at: https://www.youtube.com/watch?v=6mi7vCaIgdU&list=LL&index=2&t=1156s [Accessed 13 Jan. 2024].

Statista (2023). EU-27: GHG emissions by sector 1990-2019. [online] Statista. Available at: https://www.statista.com/statistics/1171183/ghg-emissions-sector-european-union-eu/.

Statistics Netherlands (2023). Greenhouse gas emissions 9 percent lower in 2022. [online] Statistics Netherlands. Available at: https://www.cbs.nl/en-gb/news/2023/11/greenhouse-gas-emissions-9-percent-lower-in-2022 [Accessed 17 Mar. 2024].

Steutel, F.W. and Thiemann, J.G.F. (1989). The Gamma Process and the Poisson Distribution. COSOR, 8924.
Sun, S., Guan, T., Cheng, X., Zuo, P., Gao, Y., Zuo, P. and Yin, G. (2018). Accelerated aging and degradation mechanism of LiFePO4/graphite batteries cycled at high discharge rates. 8(45), pp.25695–25703. doi:https://doi.org/10.1039/c8ra04074e.

SunPower (2013). SunPower Module 40-Year Useful Life. San Jose, CA 95134: SunPower Corporation.

Sutton, R.S. and Barto, A. (2018). Reinforcement Learning : an Introduction. Cambridge, Ma ; Lodon: The Mit Press.

Theristis, M., Venizelou, V., Makrides, G. and Georghiou, G.E. (2018). McEvoy's Handbook of photovoltaics: Fundamentals and Applications. London Academic Press, pp.671–706.

Thomas, E.V., Bloom, I., Christophersen, J.P. and Battaglia, V.S. (2008). Statistical Methodology for Predicting the Life of lithium-ion Cells via Accelerated Degradation Testing. Journal of Power Sources, 184(1), pp.312–317. doi:https://doi.org/10.1016/j.jpowsour.2008.06.017.

Trivedi, C. (2020). Game Level Design with Reinforcement Learning. [online] Medium. Available at: https://medium.com/deepgamingai/game-level-design-with-reinforcement-learning-52b02bb94954.

U.S. Energy Information Administration (2016). Photovoltaics and Electricity - U.S. Energy Information Administration (EIA). [online] Eia.gov. Available at: https://www.eia.gov/energyexplained/solar/photovoltaics-and-electricity.php.
van de Wetering, S. (2023). Energy (contracts) in the Netherlands: Everything You Should Know. [online] Vandebron. Available at: https://vandebron.nl/blog/electricity-and-gas-in-the-netherlands [Accessed 21 Jan. 2024].

van der Berg, H. (2023). Changes in the Net Metering Law Postponed until 2025. [online] www.zonnefabriek.nl. Available at: https://www.zonnefabriek.nl/en/news/changes-in-the-net-metering-law-postponed-until-2025/.

Vartiainen, E., Masson, G., Breyer, C., Moser, D. and Román Medina, E. (2019). Impact of Weighted Average Cost of capital, Capital expenditure, and Other Parameters on Future Utility-scale PV Levelised Cost of Electricity. Progress in Photovoltaics: Research and Applications, 28(6). doi:https://doi.org/10.1002/pip.3189.

Verhoeven, J. (2022). Energieprijzen Sinds 1984. [online] verhoeven272.nl. Available at: https://verhoeven272.nl/jan/energie/

prijzen.html [Accessed 14 Mar. 2024].

Vilone, G. and Longo, L. (2021). Classification of Explainable Artificial Intelligence Methods through Their Output Formats. Machine Learning and Knowledge Extraction, 3(3), pp.615–661. doi:https://doi.org/10.3390/make3030032.

von Appen, J., Braslavsky, J.H., Ward, J. and Braun, M. (2015). Sizing and Grid Impact of PV Battery Systems - a Comparative Analysis for Australia and Germany. In: 2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST). doi:https://doi.org/10.1109/sedst.2015.7315280.

Wahi, P., Konstantinou, T., Tenpierik, M.J. and Visscher, H. (2023). Lower-Temperature-Ready Renovation: An Approach to Identify the Extent of Renovation Interventions for Lower-Temperature District Heating in Existing Dutch Homes. Buildings, [online] 13(10), p.2524. doi:https://doi.org/10.3390/buildings13102524.

Walker, A., Desai, J. and Qusaibaty, A. (2020). Life-cycle Cost and Optimization of PV Systems Based on Power Duration Curve with Variable Performance Ratio and Availability. Golden, CO: National Renewable Energy Laboratory. 10.2172/1601963.

Wei, P., Xia, S., Chen, R., Qian, J., Li, C. and Jiang, X. (2020). A Deep Reinforcement Learning Based Recommender System for Occupant-Driven Energy Optimization in Commercial Buildings. IEEE Internet of Things Journal, [online] 7(7), pp.1–1. doi:https://doi.org/10.1109/JIOT.2020.2974848.

Wenham, S.R., Green, M.A., Watt, M.E., Corkish, R. and Sproul, A. (2011). Applied photovoltaics. 3rd ed. Routledge.

Weron, R. (2014). Electricity Price forecasting: a Review of the state-of-the-art with a Look into the Future. International Journal of Forecasting, [online] 30(4), pp.1030–1081. doi:https://doi.org/10.1016/j.ijforecast.2014.08.008.

Werther, M., van Hooff, W., Colijn, K. and Blokdijk, R. (2020). Next-generation Solar Power Dutch Technology for the Solar Energy Revolution. Netherlands Enterprise Agency (RVO).
Wikoff, H.M., Reese, S.B. and Reese, M.O. (2022). Embodied energy and carbon from the manufacture of cadmium telluride and silicon photovoltaics. Joule. doi:https://doi.org/10.1016/j.joule.2022.06.006.

Xu, X., Jia, Y., Xu, Y., Xu, Z., Chai, S. and Lai, C.S. (2020). A Multi-Agent Reinforcement Learning-Based Data-Driven Method for Home Energy Management. IEEE Transactions on Smart Grid, 11(4), pp.3201–3211. doi:https://doi.org/10.1109/tsg.2020.2971427.

Yang, Y. (2022). Large-scale Home Energy Management Using entropy-based Collective Multiagent Deep Reinforcement Learning. In: Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence.

Zhang, X., Biagioni, D., Cai, M., Graf, P. and Rahman, S. (2021). An Edge-Cloud Integrated Solution for Buildings Demand Response Using Reinforcement Learning. IEEE Transactions on Smart Grid, 12(1), pp.420–431. doi:https://doi.org/10.1109/tsg.2020.3014055.

Zhang, X., Jiao, Z., Zhao, C., Qu, Y., Liu, Q., Zhang, H., Tong, Y., Wang, C., Li, S., Guo, J., Zhu, Z., Yin, S. and Cui, L. (2022). Review of Land Surface Albedo: Variance Characteristics, Climate Effect and Management Strategy. Remote Sensing, 14(6), p.1382. doi:https://doi.org/10.3390/rs14061382.

Zhang, X., Li, S., He, T., Yang, B., Yu, T., Li, H., Jiang, L. and Sun, L. (2019). Memetic Reinforcement Learning Based Maximum Power Point Tracking Design for PV Systems under Partial Shading Condition. Energy, [online] 174, pp.1079–1090. doi:https://doi.org/10.1016/j.energy.2019.03.053.

Zhang, Y., Lundblad, A., Campana, P.E., Benavente, F. and Yan, J. (2017). Battery Sizing and rule-based Operation of grid-connected photovoltaic-battery system: a Case Study in Sweden. Energy Conversion and Management, 133, pp.249–263. doi:https://doi.org/10.1016/j.enconman.2016.11.060.

Zhang, Y., Zhang, Z., Yang, Q., An, D., Li, D. and Li, C. (2020). EV Charging Bidding by multi-DQN Reinforcement Learning in Electricity Auction Market. Neurocomputing, 397, pp.404–414. doi:https://doi.org/10.1016/j.neucom.2019.08.106.

Zheng, W. (2022) Deep Reinforcement Learningfor Building Control A comparative study for applying Deep ReinforcementLearning to Building Energy Management. KTH Royal Institute of Technology

Zhou, L., Zhang, Y., Lin, X., Li, C., Cai, Z. and Yang, P. (2018). Optimal Sizing of PV and BESS for a Smart Household Considering Different Price Mechanisms. IEEE Access, [online] 6, pp.41050–41059. doi:https://doi.org/10.1109/ACCESS.2018.2845900.

Zhu, J., Wu, F. and Zhao, J. (2022). An Overview of the Action Space for Deep Reinforcement Learning. In: ACAI ’21: Proceedings

of the 2021 4th International Conference on Algorithms, Computing and Artificial Intelligence. ACAI '21: Proceedings of the 2021 4th International Conference on Algorithms, Computing and Artificial Intelligence. New York, NY, United States: Association for Computing Machinery.

Ziegler, M.S. and Trancik, J.E. (2021). Re-examining Rates of lithium-ion Battery Technology Improvement and Cost Decline. Energy & Environmental Science, 14(4), pp.1635–1651. doi:https://doi.org/10.1039/d0ee02681f.