INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Maximise Available Transfer Capability in a Power Network with the support of Reinforcement Learning

Diogo Esteves Matias Lopes

Master in Data Science

Supervisors:
PhD Luís Miguel Martins Nunes, Associate Professor,
ISCTE-IUL

PhD Nuno Miguel Pinho da Silva, Invited Assistant Professor,
ISEL

September, 2024

Department of Quantitative Methods for Management and Economics

Department of Information Science and Technology

Maximise Available Transfer Capability in a Power Network with the support of Reinforcement Learning

Diogo Esteves Matias Lopes

Master in Data Science

Supervisors:
PhD Luís Miguel Martins Nunes, Associate Professor, ISCTE-IUL

PhD Nuno Miguel Pinho da Silva, Invited Assistant Professor, ISEL

September, 2024

Acknowledgment

This dissertation marks the culmination of an incredible and long journey, and I would like to express my sincere gratitude to all the people who have supported and guided me along the way.

First, I would like to thank my supervisor Professor Luís Miguel Martins Nunes and my co-supervisor Professor Nuno Miguel Pinho da Silva, for their guidance, expertise and support throughout the development of this work. Their insights have been fundamental to the successful completion of this project.

I am deeply grateful to ISCTE, where my academic journey began with my bachelor's degree and continued through my master's studies. The knowledge and experiences I have gained here have been essential in shaping this dissertation. I am grateful for the learning environment, the guidance of my professors and the lasting friendships I have built during these years.

I would also like to extend my appreciation to my friends and co-workers. Their advice, support and motivation have been essential in overcoming the many challenges faced during this research.

A special mention goes to my family, whose love, patience and belief in me has been a tremendous source of strength. Their support has been everything to me during this challenging journey.

To all of you, thank you.

This dissertation was partially financed by national funds through FCT - *Fundação para a Ciência e Tecnologia, I.P.* under the projects UIDB/04466/2020 and UIDP/04466/2020.

Resumo

Esta dissertação aborda os desafios da gestão de redes elétricas, focando-se no sistema elétrico português. Os métodos de controlo tradicionais têm de ser repensados para garantir eficiência e fiabilidade neste panorama em evolução. O principal objetivo deste estudo é implementar a Aprendizagem por Reforço Profundo (DRL) como uma estratégia de controlo inovadora para otimizar a Capacidade de Transferência Disponível (ATC), um parâmetro crucial para a segurança e estabilidade do sistema. A investigação inclui uma revisão aprofundada da literatura existente sobre métodos de controlo inteligentes, destacando o seu potencial para melhorar a eficiência e as complexidades envolvidas na sua aplicação. Embora o DRL ofereça oportunidades promissoras para otimizar a gestão do sistema de energia, as conclusões indicam que a sua aplicação não está isenta de desafios. Este trabalho fornece uma perspetiva equilibrada sobre as capacidades e limitações do DRL no contexto do sistema elétrico português e contribui com informações valiosas para futuras investigações e aplicações práticas na gestão do sistema elétrico.

Palavras chave: Aprendizagem por Reforço Profundo; Rede; Controlo Autónomo; Eletricidade; Inteligência Artificial; Capacidade de Transferência Disponível.

Abstract

This dissertation addresses the challenges of managing power networks, focusing on the Portuguese electricity system. Traditional control methods need to be rethought to ensure efficiency and reliability in this evolving landscape. The main objective of this study is to implement Deep Reinforcement Learning (DRL) as an innovative control strategy to optimise the Available Transfer Capability (ATC), a crucial parameter for system safety and stability. The research includes a thorough review of the existing literature on intelligent control methods, highlighting their potential to improve efficiency and the complexities involved in their application. While DRL offers promising opportunities for optimising power system management, the findings indicate that its implementation is not without challenges. This work provides a balanced perspective on the capabilities and limitations of DRL in the context of the Portuguese power system and contributes valuable insights for future research and practical applications in power system management.

Key words: Deep Reinforcement Learning; Network; Autonomous Control; Electricity; Artificial Intelligence; Available Transfer Capability.

Acronyms

AC - Alternating Current
AI - Artificial Intelligence
ATC - Available Transfer Capability
BCF - Base Case Flow
CBM - Capacity Benefit Margin
DC - Direct Current
DQN - Deep Q-Network
DRL - Deep Reinforcement Learning
EPRI - Electric Power Research Institute
EW - Early Warning
FCT - *Fundação para a Ciência e Tecnologia*
ICT - Information and Communication Technologies
IEEE - Institute of Electrical and Electronics Engineers
IL – Imitation Learning
KPI - Key Performance Indicators
L2RPN - Learning to Run a Power Network
LSTM - Long Short-Term Memory
MIP - Mixed-Integer Linear
MLP - Multi-Layer Perceptron
MRACPF - Modified Repeated Alternating Current Power Flow
PPO - Proximal Policy Optimization
Q – Question
R&D - Research and Development
RACPF - Repeated Alternating Current Power Flow
REN - *Redes Energéticas Nacionais*
RL - Reinforcement Learning
RNN - Recurrent Neural Network
RTE - *Reseau de Transport d'Electricite*
RTM – Required Transmission Margin
RTS - Reliability Test System
SAC - Soft Actor-Critic
SLR - Systematic Literature Review
SMAAC - Semi-Markov Afterstate Actor-Critic
TEP - Transmission Expansion Planning

TRM – Transmission Reliability Margin
TTC - Total Transfer Capability
WCCI - World Congress on Computational Intelligence

Contents

CHAPTER 1

Introduction

Electricity is a crucial resource in the modern world that most people take for granted. Still, there is a complex process that needs to happen in order to have electricity at our disposal at any second.

The conventional electrical power network connects central generating stations through a high-voltage transmission system to a distribution system that directly serves customer demand. In the past, generation stations consisted primarily of steam stations, that used fossil fuels, and hydro turbines that turned high inertia turbines to generate electricity. The energy generation was mostly centralised and located far from the end user, having to be transmitted through high voltage transmission lines and then distributed at a low voltage. However, over the past century, technological advances have led to continuous modernisation of the power grids, resulting in more decentralised power networks, increased efficiency and moving towards more environmentally friendly power sources (Henderson et al., 2017; Panda et al., 2023).

With all these new energy sources and continuous technological improvements occurring, it has become more difficult to manage all the network efficiently. Furthermore, an energy consumption increase is also expected during the next decades and energy companies need to be prepared to manage it in the best possible way. Since energy storage is very expensive and part of the energy is lost during the transmission, it is essential to find a way to manage the real-time demand, that is not constant, and the challenge is to find new techniques to help us manage power networks in an efficient and secure way (Xiao & Cao, 2020).

Artificial intelligence (AI) is on everyone's lips these days, from a simple conversation between friends to major conferences bringing together the most talented minds in the field. This topic covers a wide range of automated decision-making techniques, from simple conditional logic to sophisticated neural networks, and it has become increasingly common in data-driven industries (Kaluarachchi et al., 2021).

With the increasing decentralisation of power networks and the integration of renewable energy sources, traditional grid management techniques are no longer sufficient to handle the complexity and variability of modern power systems. The need for more intelligent and autonomous systems is critical to avoid inefficiencies, reduce energy losses and ensure the stability of the grid.

This dissertation presents the development of an autonomous topology control implementation for power networks using AI algorithms, with the aim of improving efficiency and reliability. The project specifically applies these methods to a simplified model of

the Portuguese electricity grid, testing AI-based techniques for optimizing the network's topology to better manage demand, reduce losses and enhance system performance.

The identified research questions for this dissertation are the following:

- Q1 - Is it possible to manage a simplified model of the Portuguese electricity grid autonomously while maximising ATC?
- Q2 - Which DRL agent demonstrated the best performance in managing the simplified model of the Portuguese electricity grid?

This dissertation is organised into six chapters. This chapter provides an introduction to the topic and outlines the scope of the work. Chapter 2 provides a comprehensive literature review, beginning with a discussion of the methodology used throughout the study, detailing the approaches and techniques applied, followed by a comprehensive literature review, exploring the relevant research and background knowledge. Chapter 3 focuses on the methodology, providing a comprehensive overview of all the essential components needed for implementing the training and testing process. This includes the environment, platform, actions, rules, agent and reward system. Chapter 4 explains the implementation, covering the libraries, environment, dataset specifics, backend, reward function, actions and observations, along with the model agents used. Chapter 5 presents the results and discussions for the different types of agents, "Do Nothing", PPO and SAC. Finally, Chapter 7 concludes the dissertation by summarizing the findings and suggesting potential areas for future work.

CHAPTER 2

Literature Review

2.1. Methodology

To summarise the existing research on this particular topic, it was determined that the methodology to be used would be a systematic literature review (SLR), which involves identifying, evaluating and synthesising all relevant studies that contribute to the knowledge base. This is a popular method because it helps researchers to synthesise existing evidence, identify research gaps, position new research, support or refute theoretical hypotheses and generate new hypotheses. The characteristics of conducting an SLR are (Kitchenham, 2007):

(1) Define the research question;
(2) Develop the search strategy;
(3) Documenting the search strategy;
(4) Explicit inclusion and exclusion criteria;
(5) Data extraction;
(6) Quality assessment criteria;
(7) Data presentation.

Starting from point 1, the first objective is to outline a research question in order to guide the research process and facilitate the literature review. At this stage the research question defined was "How to control autonomously the Portuguese electricity network, maximising the Available Transfer Capability(ATC)?". We could divide this question into two topics, the first being to understand the possible ways to control a power grid and the second being to understand the importance of the variable ATC.

After defining the question to approach, it is necessary to develop a search strategy (point 2), identifying the primary literature or academic databases where the search primarily takes place and determining some useful keywords to initiate the identification of primary studies. Google Scholar is the main tool chosen to perform this initial search, mainly because it provides access to a wide range of literature, including articles, theses, conference papers and books, it is easy to use and often offers free access to the literature. In terms of how the search is carried out, there are two main techniques that help to filter and focus the results. To begin a primary search, it is necessary to start by looking up broader terms related to the topic, such as "power grid" and "electric network". This helps to gain an initial understanding of the topic and its fundamental concepts. With this search, it is possible to find millions of articles related to the subjects but they do not align with the specific focus intended to be addressed. With so many articles available,

it is important to start using some Boolean operators, such as "AND", to ensure that we begin to taper towards the goal. Using a search string like "power network AND ATC", the result is literature more focused and detailed on one of our main objectives, going from millions of articles to around two hundred thousand. In this phase, one of the articles that stood out was (Pandey et al., 2010), which made me realise that an important factor to consider was how to manage the network topology.

In bullets 3 and 4, the objective is to define directives that are relevant to the goal that we want to achieve and to define which articles are selected as relevant ones. The first directive defined was to search only for documentation in English. As one of the most widely spoken languages globally, English provides access to a vast amount of literature, as many authors aim to reach the widest possible audience by publishing their work in this language. The second directive defined was to look for literature published recently, preferably in the last 5 years. This does not mean that all literature prior to 5 years ago is discarded, the objective is to give priority to the most recent because it is expected to have more updated information and innovative techniques. Additionally, these studies include in their references previous work on similar topics, which can also be reviewed if they seem relevant. The last directive defined is to give priority to studies in the field of power networks. Although there are techniques that can be used and are transversal to different areas, in the last years there has been a lot of research in this area that considers many of the latest techniques applicable in this field. Therefore, by looking at this literature it is possible to understand what has been investigated and what future work the authors could identify.

Having these guidelines in mind, the search was carried out with the query "power network AND ATC AND topology" and filtered by the years 2019 to 2023, giving a result of around 9500 articles. It is necessary to be careful when using certain queries, because if add "lang:en" to the query, the results obtained decrease to just around 40 articles, but given the title and the abstract they would not be so relevant. By reading the title and the abstract, it is possible to determine if an article is related to the main topic. However, if the research does not deep dive into the final objective of this article, it is selected only for introductory purposes. If it approaches an interesting technique or has a similar objective, it is considered eligible for further reading.

As mentioned in bullet 5, the next step is to extract the data and, to help organise all the documents and references, it was used the application Mendeley Reference Manager. It allows to create libraries, sort references into folders, and tag them for easy management and retrieval. At this stage, the document is completely extracted in PDF format and some notes are made to identify the most important topics discussed in the article, some citations that stand out and relevant information to focus more deeply in a second analysis.

The articles were selected with a focus on ensuring the highest quality and relevance to the research objectives (bullet 6). The chosen articles were subject to a quality assessment

process that included an evaluation of the robustness of their methodology, the significance of their findings and the credibility of the journals in which they were published. By prioritising articles from reputable journals, we aimed to ensure that the research included in our literature review met high standards of academic excellence and contributed meaningfully to the overall integrity and validity of our research findings.

The literature review involved an extensive search, with over one hundred titles and abstracts read to identify relevant information. However, the final selection process focused on fewer than 40 key references which form the basis of this review. Not all of these key sources were initially discovered through traditional Google Scholar searches, some were found from the citation lists within the articles considered relevant, demonstrating the importance of a versatile approach to literature review.

As identified in point 7, the SLR results are presented in the following chapter.

2.2. Related Work

As the electricity demand grows, transmission grids operate close to their thermal and voltage limits. It becomes essential to determine the maximum additional power transfer capacity between different points in the power system, encompassing transfers from generators to loads. Available Transfer Capability (ATC) quantifies the transfer capacity available beyond commitments to existing contracts or operational requirements (Chauhan et al., 2023). ATC is calculated by subtracting both the Base Case Flow (BCF) and Required Transmission Margin (RTM) from the Total Transfer Capability (TTC) (Gravener & Nwankpa, 1999). TTC is the maximum electricity quantity transmissible across a specific connection without causing thermal overloads or violating security constraints, such as voltage limits or transient stability. The RTM, which consists of the Transmission Reliability Margin (TRM) and the Capacity Benefit Margin (CBM), ensures the stability of the grid during unforeseen scenarios like generator or transmission line failures. The CBM provides excess generation capacity as a buffer against unexpected events, such as sudden load growth or unplanned generation outages.

As identified in the previous chapters, one of the objectives is to maximise the variable Available Transfer Capability (ATC). The literature underscores the crucial role of ATC in ensuring the reliable and economic facilitation of electricity transactions among market participants. Researchers emphasise the need for sufficient transmission capacity to meet the escalating demand while accommodating both renewable and non-renewable energy sources. Moreover, the importance of ATC extends beyond its role in facilitating transactions, it has a profound impact on market efficiency by influencing competitive dynamics and the reliability level of the power system. Ensuring open and non-discriminatory access to the transmission network for all market players becomes contingent on maintaining adequate ATC (Alshamrani et al., 2023).

The interest in developing a research project that addresses the topic of autonomous control of a power network was first raised through the awareness of a competition called "Learning to Run a Power Network" (L2RPN). It consists of a series of events that

challenge participants to develop artificial intelligence (AI) agents capable of operating power grids efficiently and reliably. It is organised by Reseau de Transport d'Electricite (RTE) and the Electric Power Research Institute (EPRI) with the support of various partners from academia, industry, and government.

As an introduction to the topics of electricity and power networks, the white paper (Kelly et al., 2020) provides a clear overview of the industry and it is really useful to understand the need to guarantee the stability, reliability, efficiency and adaptability of an energy network. The challenge used a simulated electricity network environment that was designed to be realistic and challenging. The paper then describes the proposed Reinforcement Learning (RL) framework for electricity network operation. The framework consists of several components, including a state space that represents the current state of the electricity network, an action space that represents the actions that the RL agent can take, a reward function that provides feedback to the RL agent on the quality of its actions and a policy function that maps from states to actions. The RL agent learns to control the electricity network by interacting with the simulated environment and receiving rewards for successful actions.

Other articles, such as (Appelrath et al., 2012; Fang et al., 2012; González Vázquez et al., 2012; Nardelli et al., 2014), were also valuable in providing context for this research. They explain that the electric power grid is responsible for delivering electricity from power plants to end users, and traditionally operates as a one-way system, where generators produce electricity that travels long distances through high-voltage transmission lines before reaching the distribution network for consumer delivery. However, the interconnected nature of its components (generation, transmission, and distribution) poses challenges, as evidenced by large-scale blackouts. In contrast, modern smart grids leverage information and communication technologies (ICT) to collect data to enhance efficiency, reliability, and sustainability. Smart grids integrate diverse energy sources and enable real-time management, presenting both technological and market-related challenges.

After understanding the fundamentals of a power network and knowing that the L2RPN competition had similar objectives to this research, articles related to the project were sought. It is important to understand how certain teams developed the project and the methodologies and techniques they used to identify trends in research approaches, highlight gaps in the existing literature and contribute to the overall credibility and trustworthiness of our literature review.

The most relevant articles were (Lan et al., 2020; Marot et al., 2020; Yoon et al., 2021). The first one, entitled "AI-Based Autonomous Line Flow Control via Topology Adjustment for Maximising Time-Series ATCs", integrated the competition in 2019 and presents an AI-driven strategy aimed at optimising the transfer capabilities of time-series data (ATCs) through autonomous topology control, taking into account various practical constraints and uncertainties.

The concept of network topology control was initially proposed in the early 1980s (Glavitsch, 1985; Mazi et al., 1986), focusing on multiple control objectives such as cost minimization, voltage regulation, and line flow control. This study addresses the challenges associated with solving the multivariate discrete programming problem inherent in transmission line switching or bus splitting/rejoining. More recent research efforts have used mixed-integer linear programming (MIP) models with DC (Direct Current) power flow approximations, employing optimisation solvers such as CPLEX (Fisher et al., 2008). (Fuller et al., 2012) introduces an approach to accelerate convergence within the described modelling and solution framework, with analogous methods detailed in (Alhazmi et al., 2019; Dehghanian et al., 2015). These last approaches incorporate point estimation techniques to model system uncertainties, using AC (Alternating Current) power flow viability checks and correction modules.

As identified in (Lan et al., 2020), while various approaches have been proposed in the existing literature, limitations persist in current methods. Relying on linear approximations in DC power flow overlooks important safety constraints and affects solution accuracy for real-world power systems, and opting for full AC power flow leads to non-convexity, presenting a challenge for effective solutions without compromising certain safety constraints or solution accuracy. In addition, the exponential growth of the simultaneous switching set for lines and bus bars increases the time required for optimisation in large power systems, impeding the real-time deployment of solutions.

The approach in (Lan et al., 2020) incorporates several AI methods, including enhanced supervised learning and deep reinforcement learning (DRL), to effectively train AI agents to achieve optimal performance. First, imitation learning (IL) is used to establish a strong foundational policy for the AI agent. Subsequently, DRL algorithms train the agent using a newly developed guided exploration technique, which significantly improves training efficiency. Lastly, an Early Warning (EW) mechanism is developed to assist the agent in identifying effective topology control strategies during extended testing periods.

Other authors concentrate on regulating electricity generation or load (Huang et al., 2020; Venkat et al., 2008; Zhao et al., 2014) but, as also mentioned in (Yoon et al., 2021), the ultimate objective is to control the power grid through topology control. This involves changing the connections of power lines and bus assignments in substations. The main goal is to reconfigure the power grid's topology, facilitating the efficient redirection of electricity flow. The approach used in this last paper (Yoon et al., 2021) led the team to win the L2RPN challenge in 2020 using an off-policy actor-critic approach that effectively tackles the unique challenges in power grid management by reinforcement learning, adopting the hierarchical policy together with the after-state representation. The off-policy actor-critic approach used was the Semi-Markov Afterstate Actor-Critic (SMAAC), a deep reinforcement learning approach that combines the after-state representation with a hierarchical decision model, demonstrated to be very effective for power grid management.

In addition to defining the method to be used to manage the electricity grid, it is necessary to define other aspects that are the basis of all the work and that could change the direction of the research. Some of these topics are well-identified in the articles (Lan et al., 2020; Marot et al., 2020) and need to be taken into account. The objective is to create a picture as close to reality as possible, making it necessary to define scenarios that introduce volatility and uncertainty into the model. This includes considerations such as daily load fluctuations, real-time adjustments, voltage set points for generator terminal buses, network maintenance schedules, and potential contingencies. All scenarios of interest should incorporate several strict constraints, that lead to failure if not met, and some soft constraints, that lead to penalties. Some constraints defined may be the same as the ones used in the competition, as they were also created with the same objective. To guide the AI agent to manage the power grid and maximise the variable ATC it is necessary to define the reward function. Taking into account these rewards in each step is also important for determining which method was the best used to manage the power grid.

Summaries of the five articles identified as the most relevant to this study are provided in Tables 1 through 5.

To conclude this literature review, although significant progress has been made in understanding and applying reinforcement learning techniques to power network control, there is still important work to be done. In particular, this dissertation attempts to apply some of the ideas from these approaches in a case simulation based on data from the Portuguese power network. The goal of this simulation is to evaluate the practical potential of these concepts, with a focus on improving grid stability and maximizing Available Transfer Capability (ATC).

Table 1. Chauhan et al., 2023

Title	A streamlined and enhanced iterative method for analysing power system available transfer capability and security
Authors & Year	Chauhan, R., Naresh, R., Kenedy, M., & Aharwar, A. (2023)
Objectives	Introduce and evaluate a novel approach, modified repeated alternating current power flow (MRACPF) with a step-size control mechanism, for analysing Available Transfer Capability (ATC) in power systems. The study compares MRACPF with existing techniques in the context of bilateral and multilateral wheeling transactions, assessing its efficiency and accuracy.
Results	The MRACPF method demonstrates comparable results to the traditional repeated alternating current power flow (RACPF) method in determining ATC. Significantly, MRACPF achieves this with a noteworthy reduction in computation time, making it a more efficient option for larger power systems.

Table 2. Alshamrani et al., 2023

Title	Transmission Expansion Planning Considering a High Share of Wind Power to Maximize Available Transfer Capability
Authors & Year	Alshamrani, A. M., El-Meligy, M. A., Sharaf, M. A. F., Mohammed Saif, W. A., & Awwad, E. M. (2023)
Objectives	Develop a methodology for enhance the Available Transfer Capability (ATC) in a transmission network by integrating wind power investment. The study adopts a bi-level structure, optimising Transmission Expansion Planning (TEP) and wind power investment at the upper level, and determining optimal generation scheduling and ATC at the lower level.
Results	The proposed method achieves an improvement in ATC on the IEEE 24-bus RTS, indicating its effectiveness in accommodating more power transactions. Although associated with increased investment costs, the enhanced ATC results in reduced load shedding and wind curtailment, highlighting the model's benefits for system security and competition among market participants.

Table 3. Kelly et al., 2020

Title	Reinforcement Learning for Electricity Network Operation
Authors & Year	Kelly, A., O'Sullivan, A., de Mars, P., & Marot, A. (2020)
Objectives	Introduce the challenge L2RPN 2020 that aims to evaluate the potential of Reinforcement Learning (RL) in optimising electrical power transmission for cost-effectiveness and safety, especially in the context of changing electricity systems. It seeks to develop RL solutions that can enhance grid security, facilitate renewable resource integration and minimise costs in real-time network operations.
Results	Reinforcement Learning (RL) is identified as a valuable framework to address evolving challenges, offering a gamified approach to developing novel solutions for network operation. The success of the L2RPN 2019 challenge underscores the efficacy of RL, leading to the launch of a larger challenge in 2020 to further advance practical RL methods for assisting power network operators in decision-making.

Table 4. Lan et al., 2020

Title	AI-Based Autonomous Line Flow Control via Topology Adjustment for Maximizing Time-Series ATCs
Authors & Year	Lan, T., Duan, J., Zhang, B., Shi, D., Wang, Z., Diao, R., & Zhang, X. (2020)
Objectives	Develop an AI-based approach to maximise time-series available transfer capabilities (ATCs) through autonomous topology control, taking into account practical constraints. Techniques including supervised learning and deep reinforcement learning (DRL) are used to train effective AI agents.
Results	Techniques such as dueling Deep Q-Network (DQN) and imitation demonstrates the AI agent's ability to effectively address the optimal topology control problem in power grids face uncertainties. Future efforts concentrate on enhancing RL agent performance, with plans to integrate the methodologies into the Grid Mind platform for autonomous grid operation and control.

Table 5. Yoon et al., 2021

Title	WINNING THE L2RPN CHALLENGE: POWER GRID MANAGEMENT VIA SEMI-MARKOV AFTERSTATE ACTOR-CRITIC
Authors & Year	Yoon, D., Hong, S., Lee, B.-J., & Kim, K.-E. (2021)
Objectives	Address the challenges of managing real-world scale power grids using reinforcement learning (RL) by presenting an off-policy actor-critic approach called SMAAC. The approach uses a hierarchical policy and afterstate representation to navigate the massive state and action space in power grid management.
Results	SMAAC proves highly effective in power grid management, ranking first in the L2RPN WCCI 2020 challenge. The approach avoids disastrous situations while maintaining high operational efficiency in diverse test scenarios, outperforming several baselines in real-world scale power grids. The work demonstrates the potential for intelligent agents to autonomously operate power grids for extended periods without expert intervention.

CHAPTER 3

Methodology

Building on the analysis of the literature from the previous chapter, the next step involves identifying and selecting the most relevant insights to integrate into the development of the model. This process is critical for ensuring that the model is grounded in the most up to date and effective approaches, drawing on key findings and methodologies from the reviewed studies.

3.1. Environment

First of all, starting with the structure of the power grid, most of the previous work has been tested and implemented in the IEEE 14-Bus System. This is a system widely used in the academic world for research, development and testing of power system studies. This system provides a manageable and sufficiently complex model for studying and understanding the dynamics of electrical power networks, which include 14 buses, 5 generators, 11 loads, 20 transmission lines and transformers, as illustrated in the following figure.

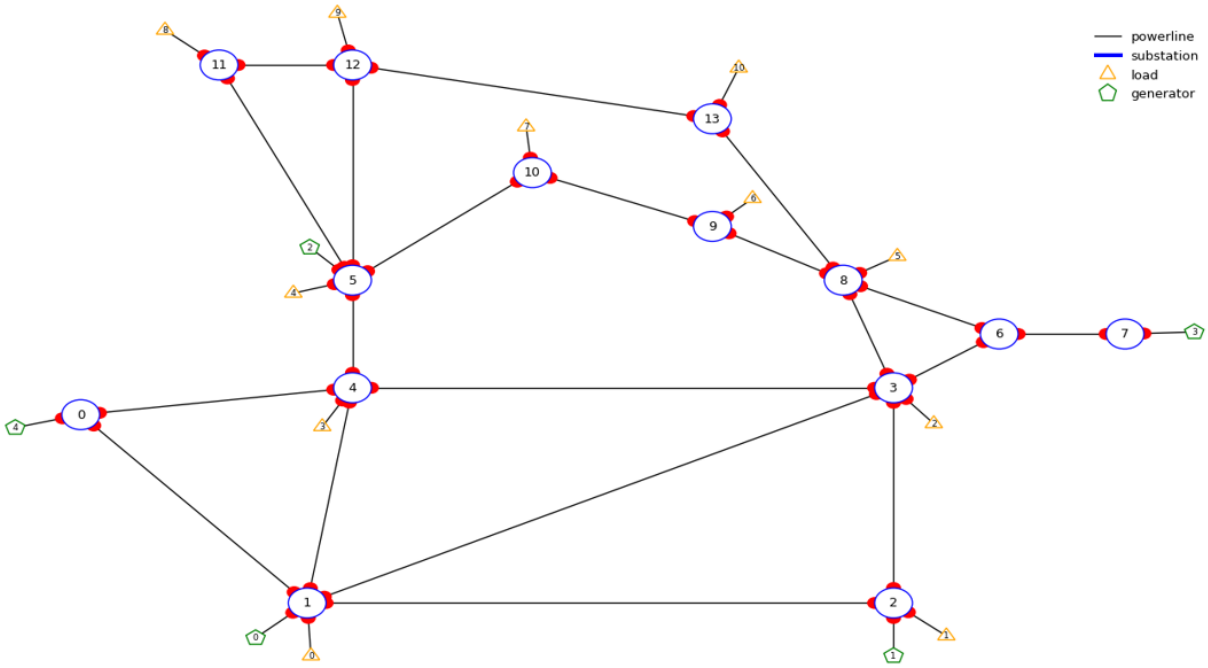

FIGURE 1. IEEE 14-Bus System

As can be seen in Figure 1, the buses/substations are represented by circles, the generator by pentagons, the loads by triangles and the transmission lines by lines connecting the buses. The generators and loads are always linked to a bus, so throughout this document, when it is necessary to identify a specific generator or load, the name referred to

consists of the number of the bus, plus an underscore ("_") and the number of the generator or load, for example, the generator 0 is linked to the bus 1 so it is called "Gen 1_0". The system consists of 20 transmission lines interconnecting various buses, with each line uniquely identified by a number. Each line connects a specific pair of buses as follows: Line 0 connects buses 0 and 1, Line 1 connects buses 0 and 4, Line 2 connects buses 1 and 2, Line 3 connects buses 1 and 3, Line 4 connects buses 1 and 4, Line 5 connects buses 2 and 3, Line 6 connects buses 3 and 4, Line 7 connects buses 5 and 10, Line 8 connects buses 5 and 11, Line 9 connects buses 5 and 12, Line 10 connects buses 8 and 9, Line 11 connects buses 8 and 13, Line 12 connects buses 9 and 10, Line 13 connects buses 11 and 12, Line 14 connects buses 12 and 13, Line 15 connects buses 3 and 6, Line 16 connects buses 3 and 8, Line 17 connects buses 4 and 5, Line 18 connects buses 6 and 7, and Line 19 connects buses 6 and 8.

Giving a brief definition of these components: buses are nodes in the system, which are points where power is either injected or withdrawn; generators provide active and reactive power to the system; loads represent the consumption of power by the end user; transmission lines connect the buses and have specific impedance characteristics that affect power flow and system stability; and transformers which step up or step down voltage levels between different buses to facilitate efficient power transfer (Kelly et al., 2020).

This work has the objective to reflect the characteristics of the Portuguese power system and one of the most important aspects is to reflect the types of energy produced. A large proportion of the energy produced in Portugal comes from renewable sources, so from the 5 generation points that we have available in this system, it was considered 1 from solar energy (Gen 7_3), 1 from wind energy (Gen 5_2), 1 from hydro energy (Gen 0_4), 1 from a mix of wind and solar (Gen 1_0) and 1 thermal energy (Gen 2_1).

Solar, wind and hydro energy are primary forms of energy derived from natural resources. Solar energy, derived from the sun's radiation, is converted into electricity using photovoltaic cells or into heat using solar thermal collectors. Wind energy captures the kinetic energy of the wind and converts it to rotational energy and then to electricity through wind turbines. Hydropower, or hydroelectric power, uses the power of moving water to generate electricity, typically through the construction of dams and the subsequent release of water through turbines (in Portugal we consider that we have hydroelectric power from reservoirs, run-of-river and small-scale hydro). Thermal energy, which includes heat energy, can come from a variety of sources and it is not exclusively classified as a renewable energy source like the others (in Portugal we consider that we have thermal energy from natural gas, biomass and cogeneration) (Ottmar Edenhofer et al., 2012; Rahman et al., 2022).

In order to apply the test to the Portuguese environment, it was used data provided by R&D Nester, a global and independent R&D (Research and Development) centre with a multicultural focus on innovation for a smarter, cleaner, more efficient and sustainable

energy system owned by REN - *Redes Energéticas Nacionais* (the Portuguese Transmission System Operator). The main data required to train and evaluate the model are the active power and voltage magnitude of generation, the active power and reactive power of consumption, and the thermal limits of the power lines. Active power is the real usable power that performs work, either being the usable power that is produced or the usable power that is consumed by homes, industries and other users. Voltage magnitude is the strength or pressure of the electricity, for example, higher voltage allows more power to be transmitted over longer distances. Reactive power is essential for maintaining voltage levels and efficient power transfer and could be described as opening and closing a valve in the pipe to regulate pressure. Thermal limit is the maximum temperature or heat that a system, component or material can withstand before it degrades or fails (Kelly et al., 2020).

3.2. Platform

In order to carry out the power grid simulations, it was necessary to select the framework to use. Primarily, because it was used in the L2RPN competition and referred to in several articles analysed in the previous chapters, the chosen framework was PyPowNet, a Python-based power system simulation framework, that provides a platform for modelling and analysing power grids. It focuses on network topology, power flow calculations and various power system components. As it was indicated by these previous studies that was a good framework to use, we started testing in this framework, but as we investigated this tool in more detail, we found another platform called Grid2Op that is considered an upgrade to PyPowNet (RTE France, 2024). Grid2Op is a reinforcement learning environment specifically designed for power grid management. It builds on the foundations of PyPowNet and differs by introducing dynamic elements, real-time decision making and a reinforcement learning framework. While PyPowNet provides a static representation of the power grid for analysis, Grid2Op creates a dynamic environment where agents must continuously adapt to changing conditions, making it a suitable platform for developing AI-driven grid management strategies.

Python is an interpreted programming language renowned for its simplicity and versatility. Its extensive standard library and the wide range of third-party modules available make it a favourite for many applications, including data science, machine learning and deep reinforcement learning (DRL). DRL, a technique that combines reinforcement learning with deep neural networks, allows agents to learn how to make optimal decisions directly from raw sensory input. Python's rich set of libraries is essential for developing and testing DRL algorithms, providing the tools and flexibility needed to advance this innovative field (Kadiyala & Kumar, 2017; Saabith et al., 2019).

Grid2Op has several features that make it work and one example that did not require any action or change on my part but without which nothing would work is the backend. The backend acts as a bridge between the actions and the grid. It translates user actions into changes that the simulator can process. The simulator performs power flow

calculations and returns the results to the backend, which then communicates them to the system. Two suitable backends were identified, "PandaPowerBackend" and "LightSim2Grid", and although they have the same purpose and many similarities, they also have some differences.

Both have been developed with the aim of simulating power grids, including real-time simulations, offering benefits such as: accuracy, ensuring that trained models can effectively generalise to real-world tasks; and smooth integration with Grid2Op, providing effective grid understanding. Regarding the differences, "PandaPowerBackend" is based on PyPower, a Python toolbox for power system analysis, which is more suitable for power flow calculations and fault analysis but may not be as flexible as "LightSim2Grid" in terms of defining custom components. "LightSim2Grid" is based on Simply, a discrete event simulation library, highly suitable for large-scale simulations and complex power system topologies but may have limitations in terms of specific functionalities compared to "PandaPowerBackend".

With the topology model defined and the necessary data prepared for the model, the base information for all tests is established. Additional components must now be selected or developed to simulate and control the power grid.

3.3. Actions

In Grid2OP several possible actions are available, that can be taken at each time step to produce changes in the power grid. These actions can be split into three categories:

- Status of the power lines: These actions are used to connect or disconnect a power line.
- Substations configuration: Substations are junctions where multiple electrical circuits meet. It is a point where electricity is distributed from one source to multiple destinations or collected from multiple sources into one line. By controlling or switching the flow of electric current by connecting or disconnecting various sections, it is possible to change the topology of the grid and manage energy flows to optimise the overall performance of the power grid.
- Injection: Change the active power of the generator production and voltage magnitude.

This research focuses on the first two options, as the objective is to study topology changes and Portugal's energy production comes primarily from renewables, so the generations provided by R&D Nester remain unchanged. In contrast, when analysing the French energy system the last option would be more relevant, given the presence of nuclear power, which allows real-time control of generation.

Regarding the action itself, the possibilities available are "set" and "change" the status. Both actions have the same objective and the only difference is that for the set action, it is necessary to provide 2 elements, the id of the object you want to modify and where you want to place it, and for the change action it is only necessary 1 element, the id of

the element you want to change. To be clearer, it was decided to use the set action, as this requires the model to identify the intended action, not just change the current state.

For this simulation, the action values used represent switches in the power grid. This reflects real-world operations where switches are activated to isolate lines and potentially reconfigure the network topology.

3.4. Rules

This work aims to reproduce as closely as possible how a power grid works, so it is necessary to define several constraints so that it is not possible to perform actions that are not possible in the real world.

The following constraints are those that must be met at all times because failure to do so results in a Game Over, meaning that the simulation ends.

The first is that the system must consistently meet user demands and provide uninterrupted power. This means that users should always have access to electricity when they need it, without experiencing power outages or failures. While our simulation assumes a defined supply and consumption of energy, the real Portuguese scenario may require the purchase of electricity from Spain in real time to meet demand. The data provided by R&D Nester recognises this potential constraint. The data provided by R&D Nester ensures a simulated environment in which energy production always equals or exceeds consumption, eliminating the need for external power purchases.

The second constraint is that no more than one power plant is allowed to experience an unexpected shutdown or more than one power line can be open in each time step. Given the small size of the IEEE 14 grid, with limited number of generators and transmission lines, the loss of more than one of either would inevitably result in insufficient power generation or line capacity to meet demand. Moreover, even with a single generator or line failure, we are unlikely to be able to maintain a consistent supply of electricity.

The third constraint is the prevention of electrical island formation through topology control. An electrical island is a part of the electrical grid that is disconnected from the main grid, which can lead to power outages and instability. This constraint ensures that changes to the configuration of the power grid made by switching operations do not result in isolated sections of the grid.

The fourth and final constraint is that AC power flow should converge at all times. The mathematical representation of the power system must always converge to a stable balance where electricity flows smoothly. Convergence failures indicate potential system problems such as overloads or instabilities.

3.5. Agent

One of the most important components that needs to be defined is the types of agents tested and compared. The agent is the algorithm that is trained and used to control the actions taken during the simulation.

The best algorithms to use depend on the specific problems and environment of the project. Three basic concepts that need to be known are the differences between:

- On-Policy vs. Off-Policy: On-policy algorithms are typically faster to learn but require interaction with the environment during training. Off-policy algorithms can learn from pre-collected data but may require more exploration strategies (Fakoor et al., 2020).
- Discrete vs. Continuous Actions: Some algorithms, such as Deep Q-Network (DQN), are better suited to discrete action spaces, where choices are limited (e.g. "left", "right"). Others, such as actor-critic methods (e.g. PPO, SAC), handle as well continuous actions, where actions can take any value within a given range (e.g. temperature).
- Exploration vs. Exploitation: Some algorithms, like SAC, prioritize exploration, meaning they actively seek out new actions to discover potential rewards. This exploration helps them find optimal solutions. In contrast, algorithms like PPO focus more on exploitation, choosing actions that are believed to give the highest reward based on past experience. Finding the right balance between exploration and exploitation is essential for successful reinforcement learning.

Based on the information gathered, it was decided that would be created agents based on three models – "Do Nothing", PPO and SAC. Each has distinct characteristics that provide valuable insights into different reinforcement learning approaches. Below is an explanation of the main characteristics of each model agent.

The "Do Nothing" agent is the most basic agent that could be used. As the name suggests, this agent does not produce any actions during the simulation. By taking no actions, it establishes the minimum level of performance, helping to assess whether more complex algorithms actually lead to improvements.

Proximal Policy Optimisation (PPO) is an on-policy reinforcement learning algorithm that stands out for its simplicity and efficiency. It belongs to the family of policy gradient methods, which means that it learns the policy (a mapping from states to actions) by optimising the expected cumulative reward. PPO aims to strike a balance between exploration (discovering new strategies) and exploitation (using known strategies) by adapting its actions based on real-time interactions with the environment. This makes it particularly well suited to dynamic and continuously changing environments, and this balance is crucial in reinforcement learning tasks like grid control, where the agent must adapt to changing conditions while optimising grid stability and performance.

PPO introduces a unique mechanism to ensure that policy updates are controlled. Rather than making large, unpredictable updates to the policy, PPO uses a clipping technique that restricts changes to a defined range. This clipping mechanism ensures that the policy does not deviate too far from the previous one during training, preventing instability and excessive changes in behaviour. As a result, PPO achieves a stable and efficient

learning process, enabling better performance in a variety of environments (Engstrom et al., 2019; Schulman et al., 2017).

Soft Actor-Critic (SAC) is an off-policy reinforcement learning algorithm, which means it can learn from interactions with the environment as well as from past experience. SAC's main strength lies in its focus on exploration. By encouraging the agent to explore a wider range of actions, SAC avoids getting stuck in local optimal and increases the chances of finding globally optimal strategies.

The central idea behind SAC is the concept of maximum entropy reinforcement learning. Unlike traditional algorithms that only aim to maximise rewards, SAC also maximises entropy, which encourages the agent to choose different actions. This results in more robust policies and ensures that the policy remains flexible and exploratory, helping the agent to deal with unexpected situations (Haarnoja et al., 2018).

3.6. Reward

For the agent to have good performance, a crucial aspect that needs to be defined is the reward function. It directly influences how the agent learns to make decisions and achieve the goal of maximizing it.

As already mentioned, our goal is to maximise the ATC and to do so it is necessary to define the reward function to produce these results.

ATC (Available Transfer Capability) assesses how effectively an agent manages the flow of current across power lines to ensure that the system operates safely within its limits. ATC measures the amount of additional power that can be transferred through the network while maintaining safe margins below the thermal limits of each power line. The reward is based on the margin for each power line, which reflects how close the current flow is to the thermal limit, the maximum current the line can carry without overheating. In this way, the performance of the agent is directly linked to maximising ATC by keeping power flows within safe limits.

The margin for each power line is calculated as the difference between the thermal limit and the actual current flow, divided by the thermal limit. This ratio indicates how close the power line is to reaching its critical threshold. However, instead of using this basic ratio, the margin is calculated as one minus the square of the relative flow, where the relative flow is the ratio of the actual current flow to the thermal limit. Squaring the relative flow increases the penalty for smaller margins, meaning the reward function becomes more sensitive to lines operating close to their thermal limits. This ensures that the reward decreases more as a power line approaches its thermal limit, encouraging the system to avoid risky states where lines are heavily loaded.

If the current flow is less than or equal to the thermal limit, the margin is a positive value between 0 and 1, where higher values indicate safer operating conditions with more unused capacity. However, if the current flow exceeds the thermal limit, the margin is set to zero, representing an unsafe condition where the line is overloaded.

The reward function then takes this margin for each power line and sums these margins over all the power lines in the system. As a result, the reward is higher when all power lines are operating well below their thermal limits, reflecting a safer and more stable grid. Finally, the final reward is the cumulative reward across all time steps processed.

Following is the mathematical formula of this reward function:

$$\text{(1) Power Line Margin} = \begin{cases} 0, & \text{if } Flow \geq ThermalLimit \\ 1 - \left(\frac{\text{lineflow}}{\text{thermallimit}}\right)^2, & \text{otherwise} \end{cases}$$

$$\text{(2) Time Step Reward} = \sum_{i=1}^{nlines} PowerLineMargin_i$$

$$\text{(3) Total Reward} = \sum_{j=1}^{ntimesteps} TimeStepReward_j$$

In practice, by choosing this reward function, the agent is incentivised to manage power flows in such a way that all power lines maintain a large margin by staying well within their safe operating limits. By maximising this reward, the agent learns to distribute power flows in a way that minimises the risk of overloading any individual power line. This approach is particularly important in real-world applications where the integrity and reliability of the power grid are critical.

CHAPTER 4

Implementation

In this chapter, after describing the major components present in Grid2Op, it was time to start implementing the entire process that has been idealised. With all the information gathered and analysed we have at our disposal all the necessary information to implement the simulation of the Portuguese power grid.

At this stage, it was necessary to define the components used to implement this work: the libraries, the environment of the power grid, the dataset, the backend, the reward function, actions and observations.

4.1. Libraries

The first step to do is to import all the necessary libraries. Since we are using Grid2Op, the first library to import was "grid2op", but to use directly the names of the classes in several cases we import directly the classes from the submodule, as shown in the following example:

```
from grid2op.Reward import L2RPNReward
reward_class=L2RPNReward
```

Regarding the backends, the "PandaPowerBackend" is part of "grid2op.Backend", so it is imported as referred above, but to use the "LightSimBackend" it is necessary to import the library "lightsim2grid", class "LightSimBackend".

The models used by the agents to learn and control the power grid were imported from the library "stable_baselines3".

This work also uses the "gymnasium" library, which is discussed in more detail in this chapter, along with supporting libraries such as "shutil," "tqdm," "os," "collections," "IPython," and "matplotlib".

4.2. Environment

As referred to in the last chapter, the objective is to implement the simulation in the IEEE14 grid, composed of 14 substations. It was necessary to build the grid adapted to the Portuguese environment, which could be divided into 3 main pillars: the representation (visualization) and components of the grid, the capability of each component and data in each time step (e.g.: production and consumption).

The IEEE 14-bus system is a widely used grid model due to its versatility. To represent the visual structure and components of a power grid, we adopted a layout from previous test environments used in the L2RPN competition and present in Grid2Op. The

base structure was derived from the environment called "l2rpn_case14_sandbox" and its representation is shown in Figure 1 in the previous chapter.

The only change performed was to change the generator 0 present in substation 1 from nuclear to a mix of wind and solar energy. The environment used as a baseline was inspired by the French power grid and nuclear power is a significant portion of the energy produced by them. On the other hand, Portugal relies on renewable energy and, at this moment, does not explore nuclear power.

4.3. Dataset

The specific details and data of each component in the grid (power lines, generators, and loads), provided by R&D Nester, were identified and analysed at this time.

In terms of data volume, a total of 2976 data points were received with a time interval of 15 minutes between each time step, which means we have 31 days of information. Given the amount of data we were able to gather, it was decided that the dataset would only be split in train and test, and skip the validation. Although validation is generally important in machine learning to assess the performance of the model on unseen data and prevent overfitting, it was considered that with the limited data at our disposal would be more beneficial to have more data to train and with that improve the effectiveness of the model. It was therefore decided to test the model with data from a full week, 7 days, and to train the model with the remaining 24 days (more or less a standard 20/80 split).

To ensure the validity of our results, the following analysis of the data provided was carried out to assess the quality and potential issues that could affect our analysis.

4.3.1. Power Lines

Regarding the power lines the only necessary thing to be defined is their thermal limit in amps, that is the maximum temperature that can be safely handled by the power line without compromising their integrity and performance. The Figure 2 shows the power grid with the identification of the thermal limits of each power line, which have not been modified from the French model as they are very similar.

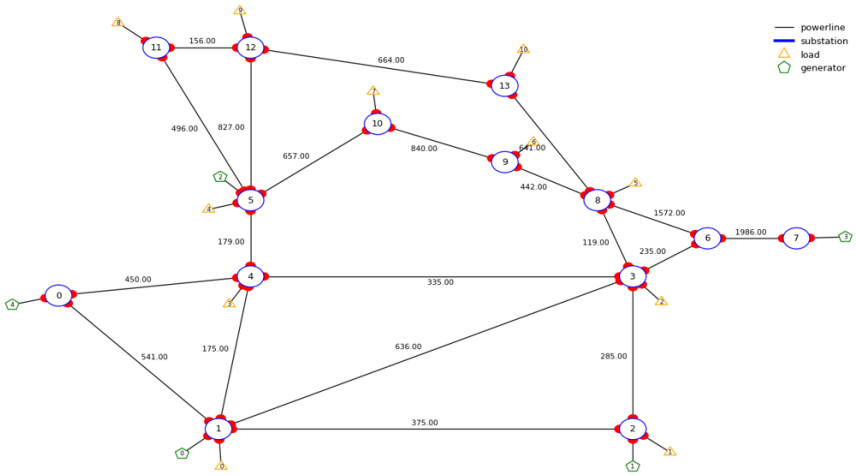

FIGURE 2. IEEE 14-Bus System with Power Line Thermal Limits

4.3.2. Generators

Regarding the generators, it is necessary to define the voltage magnitude (Table 6) and the active power of each generator. The voltage magnitude defines the strength of the electricity, so in order to ensure the best system performance, the voltage values assigned to the generators are carefully selected based on criteria such as efficiency, safety and system stability, and are set to comply with the nominal voltages of the grid. This ensures compatibility across the grid, prevents operational issues and maintains a reliable and stable electricity supply.

TABLE 6. Generators Voltage Magnitude (kilovolts)

Gen 0_4	Gen 1_0	Gen 2_1	Gen 5_2	Gen 7_3
142,1	142,1	142,1	22	13,2

As each type of generator produces energy from different sources and with a different distribution during the day, a special care was required on the part of R&D Nester in collecting the information regarding the active power for the generators, and in this step, it was time to check that the graphs of each type of generator corresponded to reality.

To perform this analysis, it was collected the data from 2 random days regarding the 4 different types of generators and created line graphs to visualize the energy distribution during the days.

For the solar energy, it was used Gen 7_3 and as it is possible to see with the help of Figure 3, as expected due to hours of sun, the graph has a normal distribution, starting to produce energy from the 8 hours and ending at 18 hours, with the peak in the middle of the day at around 12 hours. Although there may be days with more or less cloud cover affecting energy production, this generator is expected to have the most consistent behavior in terms of the hours of the day it produces energy.

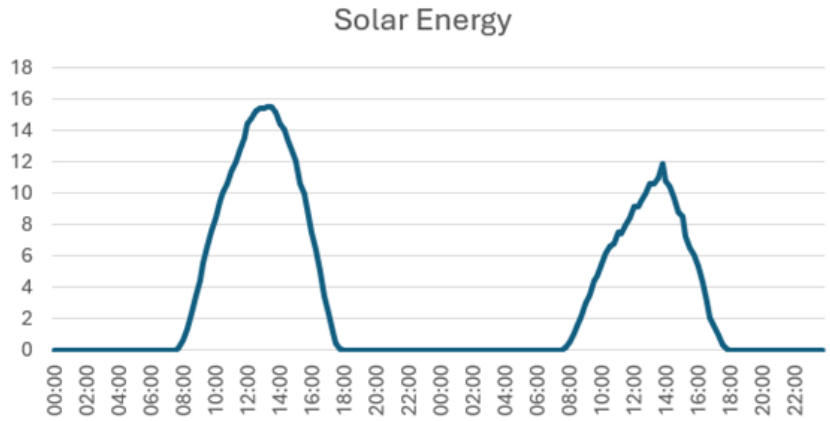

FIGURE 3. Solar Energy (megawatts)

Unlike solar energy, wind energy could be more inconsistent. We could have days with more or less wind and on the same day, we could have moments more windy than others.

Figure 4 shows the example of Gen 5_2, where it is possible to see that on the first day a lot of energy was produced until the end of the afternoon, but from then on it started to decrease, with only a small increase in the last hours of the second day.

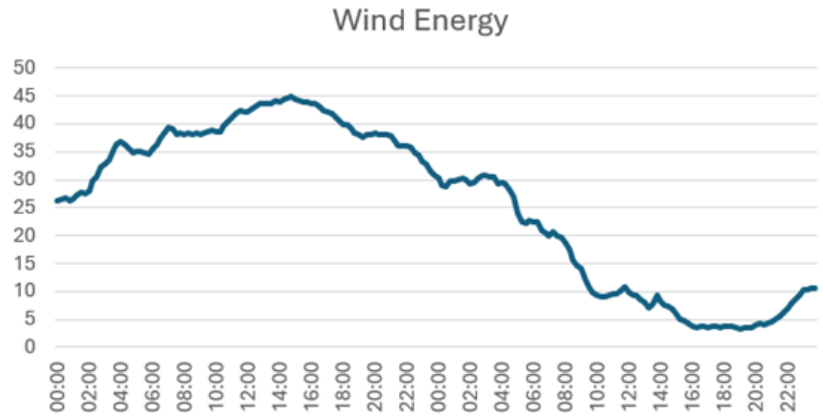

FIGURE 4. Wind Energy (megawatts)

The previous two types of generators are completely dependent on natural resources and cannot be controlled. Hydro and thermal power, on the other hand, can control some of the energy they produce.

Starting with hydro power, Portugal has built a good system of reservoirs, so it can generate more electricity when it is needed. Typically, the beginning of the working day, when people wake up and get ready, and the end of the afternoon, when everyone goes home, are the peaks of demand and the hydro power plants can quickly meet this demand by adjusting their water flow rates, as can be seen in the following graph of Gen 0_4 (Figure 5). Analysing together the energy produced by solar and hydro energy, it is also possible to conclude that they complement each other: hydro power starts early in the morning to produce a high amount of energy and as more solar energy is produced, the hydro power starts to decrease, and at the end of the day the opposite effect occurs.

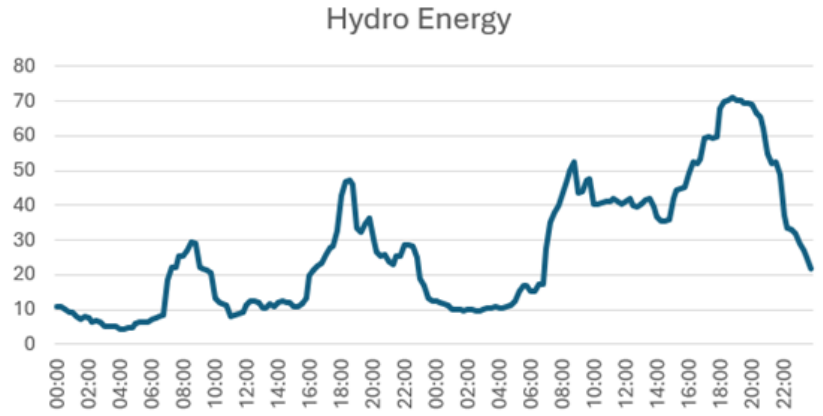

FIGURE 5. Hydro Energy (megawatts)

Regarding thermal energy, generally, the electricity generated is constant, but when there is an unexpected demand for energy or a few amount of energy is produced by the

remaining sources of energy, there is the possibility to increase the production of thermal energy, as it happens at the end of the second day of the following example of the Gen 2_1 (Figure 6).

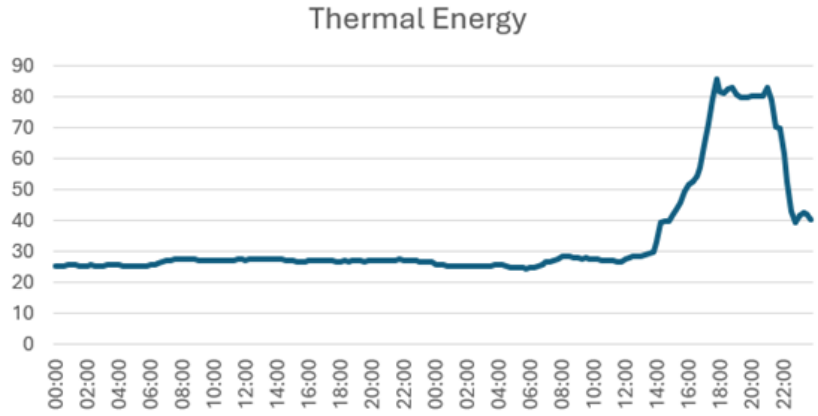

FIGURE 6. Thermal Energy (megawatts)

4.3.3. Loads

For the loads, it was necessary to define the reactive power and active power of each load. These two variables were provided by R&D Nester and they take into account all the variables previously defined as well as the objective of this project. In this work, the objective is to manage the topology of the grid without modifying the energy generated at any given time, so the major conditions in the data that they provided us were that the total energy consumed could not exceed the total energy generated at each time step.

4.4. Backend and Reward Function

These two variables have already been identified and explained in the previous chapter, so to define the environment it is only necessary to call them. For the backend, both "PandaPowerBackend" and "LightSimBackend" were tested to understand if the differences that they have produce different results. The reward function used is the one defined to maximise the ATC.

4.5. Actions and Observations

Grid2Op is a platform designed to simulate sequential decisions in power systems and provide different reinforcement learning algorithms and grid control strategies, but Gym is a widely used toolkit for developing reinforcement learning algorithms that provide a standardised interface that simplifies integration with different algorithms and libraries. To get the best out of both tools, there is already a submodule of Grid2Op that creates a Gym compatible environment to use the range of reinforcement learning algorithms already developed for Gym environments. Gym environments can be easily integrated with other tools and libraries, such as TensorFlow or PyTorch, which are commonly used

by other researchers in the field of deep learning, which could help speed up your research and development process as more information is available.

TensorFlow and PyTorch are two popular open sources used with deep learning frameworks. TensorFlow provides flexible tools for building machine learning models, particularly neural networks, and is known for its features ready to implement (Dillon et al., 2017). PyTorch is valued for its dynamic graph, which provides a more intuitive and flexible approach, especially in research and experimentation (Paszke et al., 2017).

When converting the environment to Gym, it is time to select the attributes that feed the model regarding the action and observation space.

As mentioned previously, the actions that can be taken by the agent are only topological and it was decided to use the set buses and the set line status. These options allow control over how the buses are connected and how the energy flows.

Although the observation space was not previously mentioned, it plays a critical role in building a successful model. A careful selection of the most relevant variables is essential, as including unnecessary information can mislead the model, causing it to waste time and resources trying to establish connections between irrelevant variables. By focusing only on pertinent data, we ensure the efficiency and effectiveness of the model in delivering accurate results. From a large set of 46 variables available to be selected, were chosen the following 21:

- Time variables: "day_of_week", "hour_of_day", "minute_of_hour", "time_before_cooldown_line", "time_before_cooldown_sub", "timestep_overflow" (the last three represent the number of time steps before a power line or substation becomes unavailable due to a "cooldown" period, and the number of time steps since a power line has been in an overflow state).
- Input data variables: "gen_p", "gen_q", "gen_v", "load_p", "load_q", "load_v" (these variables are related with the generators and loads, and "_p" represents the active power, "_q" represents the reactive power and "_v" represents the voltage of the bus to which is connected).
- Output data variables: "p_or", "q_or", "v_or", "p_ex", "q_ex", "v_ex" ("_or" represents the origin side and "_ex" the extremity side of the power lines, giving the active power, reactive power and voltage).
- Topology status variables: "line_status", "topo_vect" (for each load, generator and end of a power line it gives on which bus this object is connected in its substation).
- Calculated variables: "rho" (capacity of each power line, this means the observed current flow divided by the thermal limit of each power line).

4.6. Model Agents

As mentioned in the previous chapter, three different agents were modelled to be implemented: "Do Nothing", PPO and SAC.

To implement the "Do Nothing" agent, there is no need to define any parameters or settings manually. This agent is already predefined within the Grid2Op framework. It is just necessary to call the "Do Nothing" agent directly from Grid2Op without any additional configuration. It acts by taking no action during the simulation, which is useful for testing and baseline comparisons.

To implement the PPO and SAC, it was used the "stable_baseline3" which is a reliable Python library of reinforcement learning algorithms in PyTorch. To create the models, it is necessary to define several parameters that lead to different learning and actions. The first is the environment, which uses the Gym environment created earlier. Regarding the other arguments, one approach uses the proposed hyperparameters in the "stable_baselines3" GitHub repository as a baseline (Table 7), while other variations incorporate modifications that we believe could improve performance.

TABLE 7. Baseline Hyperparameters

learning_rate	1e-3
policy_kwargs	"net_arch": [200, 200, 200]
batch_size	8

Starting to explain what each parameter means: the "learning rate" is a hyperparameter that controls how much weights are adjusted concerning the loss gradient during each training iteration; the "policy" is the function that the agent uses to decide its actions based on the observed state (defined below); "policy_kwargs" is an additional argument to customize the architecture; "batch_size" is the number of training samples processed together in a single iteration before the weights of the model are updated; "verbose" only controls the level of detail in the training output, with True providing more detailed logs and False providing less; "seed" is used to ensure the reproducibility and consistency of random processes, enabling reliable comparison and debugging of experiments. Additionally, in PPO it is possible to add the n_epochs, which represents the number of times the model updates its parameters by optimizing the surrogate loss function using the same batch of data. The surrogate loss function is used to optimize the policy while ensuring that the new policy does not deviate too much from the old one.

Moving on to analyse the changes to be tested against the proposed baseline, the base policy used is the "MlpPolicy", which stands for Multi-Layer Perceptron Policy and it is a type of policy network in reinforcement learning that uses fully connected layers, also known as dense layers. This policy is effective at learning patterns and is commonly used when the input data consists of structured and non-sequential data. Has we are dealing with a temporal structure we also should use a policy based on a Recurrent Neural Network (RNN) architecture. SAC, which is typically used with non-recurrent policies due to its off-policy nature, uses standard feedforward policies like "MlpPolicy" but for PPO the implementation also includes testing a policy based on RNN. It was chosen to test an LSTM-based policy (Long Short-Term Memory), "MlpLstmPolicy", that combines a multilayer perceptron (MLP) with an LSTM layer to handle sequential

dependencies in the data. It is well-suited for environments where the agent needs to remember information from previous time steps to make decisions.

For the learning rate, besides the learning rate 1e-3, the experimented also includes a lower rate of 1e-4; for the batch size, since 8 is relatively small, the testing also incorporate 16 and potentially 32, if expected to produce a better result, because the use of a larger batch size aims to produce a more stable estimate; for the network architecture the suggestion is to use three hidden layers with 200 neurons each and the test also includes an experiment with only two hidden layers because in some cases adding more layers does not significantly improve model performance and may even lead to worse outcomes if the added complexity introduces noise or instability; for the n_epochs, the initial testing is conducted without specifying the parameter, using the default value of 10, and if the model demonstrates a need for more iterations over the same data, the parameters are fine-tuned and the process is repeated.

After defining all the different approaches to be implemented, the next step is to set up the training process. The training consists of 2304 steps, and the testing is evaluated in 672 steps, which represents one week of data.

CHAPTER 5

Results and Discussions

Starting to implement the idealised training options, a continuous evaluation is done immediately to eliminate options that might look good on paper but would not produce good results. This streamlines our training by eliminating unnecessary testing of options that we can predict would not be effective.

The testing and evaluation is conducted using two distinct scenarios. The first scenario serves as a baseline, where the agent operates under normal conditions without any additional challenges. In the second scenario, a critical event is introduced by forcing the disconnection of a crucial power line at time step 0. This approach enables a comprehensive evaluation of the agent's behaviour, decision making, reliability and network management capabilities under normal and fault conditions.

The evaluation is carried out for the three types of agents: "Do Nothing", PPO and SAC. Key Performance Indicators (KPIs) used for comparison include survival rate (completion of all steps), cumulative reward, diversity of valid actions taken, deviation from the original stable grid configuration, and number of line reconnections and disconnections during the test period. It is important to note that a power grid can have multiple valid topologies that achieve similar or better results than the original configuration. The agents may discover alternative topologies that enhance grid stability, minimise losses or manage contingencies more effectively, demonstrating their potential for innovative grid management strategies.

The Grid2Viz web application serves as a valuable tool to provide a visual understanding of the results generated by Reinforcement Learning agents running on the Grid2Op platform. By providing a visual understanding of agent behaviour and performance, it is possible to make more informed decisions about agent deployment and optimisation, and to accelerate agent refinement.

5.1. Scenario 1 - Baseline

5.1.1. Do Nothing

The "Do Nothing" agent does not require any extensive hyperparameter tuning. For this baseline agent, the only comparison to be made is between the two backend systems: "PandaPowerBackend" and "LightSimBackend".

Metrics from Grid2Viz reveal that the power line's thermal limit is nearly fully utilised at its peak but is never exceeded, as shown in Figure 7. This indicates that the system remains operational and stable in both backends even without intervention.

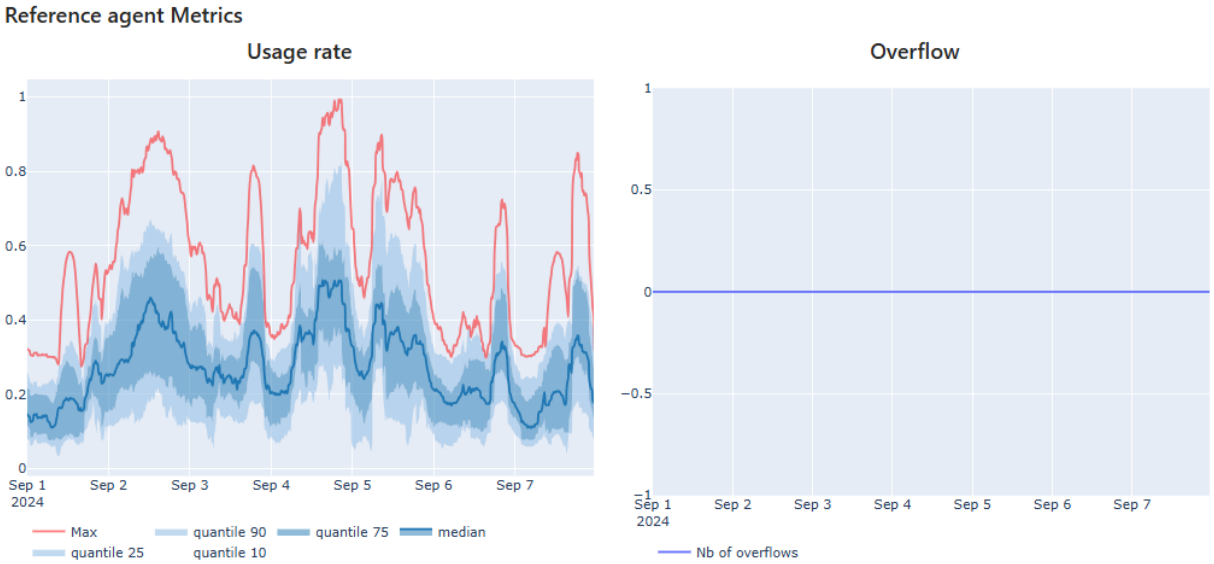

FIGURE 7. Usage Rate and Overflow

5.1.2. PPO

The testing and evaluation of the PPO agent start with four scenarios, two policy types "MlpPolicy" and "MlpLstmPolicy" across the two different backends tested with the "Do Nothing" agent. This assessment is performed using the hyperparameters proposed as baseline.

Starting with "MlpPolicy", both backends survived achieving the same reward of 11948 but with different behaviours. The agent using "PandaPowerBackend" indicates that it executed a single action that changed the grid topology by 11 units compared to the reference. The one using "LightSimBackend", while also performing a single action, maintained a consistent deviation of 9 units from the reference throughout the simulation.

Regarding the "MlpLstmPolicy", both backends demonstrated similar results, with the agents successfully surviving, achieving a reward of 11948, maintaining a grid deviation within 13 units from the reference and only performing a single action during the first time step.

Looking at these KPIs, it was initially intriguing that a single action could lead to such a significant deviation from the reference grid. Analysis of the agents' actions revealed that they were attempting to disconnect multiple lines in the single action, ranging from 9 to 13 lines depending on the agent. Although such a number of disconnections would likely result in a system failure due to insufficient capacity, Grid2Viz confirmed that these actions were not being carried out.

Further investigation revealed that the agents were proposing ambiguous or illegal actions. The rules described in chapter 3.4. clearly state that only one action is allowed per time step and that disconnecting multiple lines could create isolated sections of the grid (islands), which is also not permitted.

To address this issue, the agent was modified to comply with the rules, specifically limiting actions to a single connect or disconnect per time step. In addition, the agent

was prioritised to reconnect lines whenever possible, rather than disconnect them, and some controls were added to understand whether the actions were actually being taken.

After the agent's modifications, the same four tests were conducted, resulting in significantly different outcomes. None of the four agents were able to survive the simulations. Each agent chose to disconnect one line per time step, but the agent using "LightSimBackend" and "MLPPolicy" disconnected less critical lines and survived for 6 time steps. The remaining agents only survived 2 time steps. Although the differences are not huge, subsequent testing of PPO uses backend "LightSimBackend" based on these results.

To assess the effect of batch size, both policies were tested using "LightSimBackend" while keeping all other hyperparameters unchanged. When the batch size was increased to 16, both agents survived, but achieved slightly different rewards. The agent using "MlpLstmPolicy" remained passive, while the "MlpPolicy" agent disconnected a line (line 0) in the first step. Although the thermal limit was not exceeded, the reward was lower due to the reduced capacity of the connected lines. When the batch size was increased to 32, both agents took no action, resulting in them completing all test steps with a reward of 11948.

Finally, an alternative network architecture with two hidden layers of 200 neurons each was evaluated for both policies. Neither agent survived the entire test, with the "MlpPolicy" agent surviving 115 time steps and disconnecting 4 power lines. The "MlpLstmPolicy" agent survived only 6 time steps, disconnecting more lines, including more critical ones.

5.1.3. SAC

SAC agents were tested and evaluated using the same methodology as the corrected PPO agent implemented in the previous subsection. Every configuration used with PPO was replicated for SAC and the results were equal across all tests. Each configuration remained stable, achieving a reward of 11948 without taking any active steps, maintaining the same topology as the reference network.

Although these baseline results demonstrate stability, they highlight the need for further exploration. The upcoming tests are critical in determining whether the SAC agent can go beyond maintaining the current grid state and adapt dynamically to changing grid demands, adjusting its actions accordingly.

Tables 8 through 10 present a summary of the key results of the baseline scenario for each type of agent.

TABLE 8. Do Nothing Agents Results

Do Nothing

Backend	Survival	Reward	Distinct Actions
PandaPowerBackend	Yes	11948	0
LightSimBackend	Yes	11948	0

Table 9. PPO Agents Results

PPO

Backend	Policy	Learning Rate	Batch Size	Policy_kwargs	Survival	Reward	Distinct Actions	Topology from	Average Disconnections per Action
PandaPowerBackend	MlpPolicy	1.00E-03	8	[200, 200, 200]	2	19	2	1-2	1
LightSimBackend	MlpPolicy	1.00E-03	8	[200, 200, 200]	6	95	6	1-6	1
PandaPowerBackend	MlpLstmPolicy	1.00E-03	8	[200, 200, 200]	2	19	2	1-2	1
LightSimBackend	MlpLstmPolicy	1.00E-03	8	[200, 200, 200]	2	19	2	1-2	1
LightSimBackend	MlpPolicy	1.00E-03	16	[200, 200, 200]	Yes	11888	1	0-1	1
LightSimBackend	MlpLstmPolicy	1.00E-03	16	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-03	32	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpLstmPolicy	1.00E-03	32	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-03	8	[200, 200]	115	2093	4	1-4	1
LightSimBackend	MlpLstmPolicy	1.00E-03	8	[200, 200]	6	96	6	1-6	1

Table 10. SAC Agents Results

SAC

Backend	Policy	Learning Rate	Batch Size	Policy_kwargs	Survival	Reward	Distinct Actions	Topology from Reference	Average Disconnections per Action
PandaPowerBackend	MlpPolicy	1.00E-03	8	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-03	8	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-04	8	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-03	16	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-03	32	[200, 200, 200]	Yes	11948	0	0	0
LightSimBackend	MlpPolicy	1.00E-03	8	[200, 200]	Yes	11948	0	0	0

5.2. Scenario 2 - Disconnection

In this second scenario, as mentioned at the beginning of the chapter, a line disconnection is introduced to observe the behaviour of the agents. Line 17, which connects buses 4 and 5, was selected for the disconnection because of its critical role in connecting the upper and lower parts of the grid, making it a key component in maintaining overall grid stability.

The tests performed in this section focus exclusively on those configurations that successfully remained stable over all time steps in the first scenario, with the aim of understanding whether the agents can effectively adapt to such perturbations while maintaining optimal grid performance.

5.2.1. Do Nothing

Before initiating the test for the “Do Nothing” agent with one line disconnected, it was already known that the thermal limit of the lines was close to full capacity at their peak when all lines were operational.

Given the reduced thermal capacity available from the power lines at the peak and the disconnection of one line, it was expected that the simulation would eventually become unstable. This instability occurred when the system exceeded the thermal limits at time step 113, resulting in Game Over in both backends.

5.2.2. PPO

The performance of the PPO agent was tested using the four agents configuration that successfully completed all time steps in scenario 1. These agents had in common the backend ”LightSimBackend”, a learning rate of 1.00E-03 and three hidden layers of 200

neurons, and were tested for the combination of the policies "MlpPolicy" and "MlpLstmPolicy" and the batch sizes 16 and 32.

Although the simulation in Scenario 2 started with the disconnection of a line, these agent configurations showed exactly the same behaviour as in Scenario 1. Despite the expectation that the agents would recognise the need to reconnect line 17, they all failed to take this action, resulting in its survival for only 113 time steps.

The results of the PPO agents in this scenario produced unsatisfactory results as the model struggled to make appropriate decisions. To address this, further testing was carried out using the "MlpLstmPolicy" policy, batch size 16 and increasing the number of training epochs with the objective of achieve better optimisation. The default value used was 10 epochs and it was necessary to increase this value to 500 epochs for the PPO model to be able to complete all the test steps. The agent reconnects line 17 at time step 1 and disconnects line 0 at time step 2, but as seen in the previous scenario, even with line 0 disconnected, the agent is able to survive.

This improvement probably occurred because PPO, as a policy gradient method, requires sufficient training time for the policy to stabilise and effectively learn from the environment. With more epochs, the model had more opportunities to improve its policy, explore the state-action space, and learn better strategies for managing the grid topology. The earlier poor performance may have been due to inadequate training, where the agent had not fully converged or was stuck in suboptimal policies due to insufficient training iterations.

5.2.3. SAC

Given the successful survival of all SAC agent configurations in Scenario 1, they are now evaluated in the new scenario. To initiate the SAC agent evaluation, a comparative analysis of "PandaPowerBackend" and "LightSimBackend" was conducted using the baseline hyperparameters.

After evaluating both backends, three KPIs showed identical results: both systems survived, performed a single action and reconnected line 17. However, the timing of this action was different, the "LightSimBackend" reconnected line 17 at time step 3, while the "PandaPowerBackend" delayed reconnection until time step 35, resulting in a slightly lower reward in the latter case. Although these differences were not substantial, "LightSimBackend" was chosen for further testing due to its slightly more efficient performance.

The test was carried out by modifying the learning rate to 1e-4 while using "LightSimBackend" and baseline hyperparameters. The agent chose not to reconnect any lines and this decision led to its survival for only 113 time steps, mirroring the outcome of the "Do Nothing" agents.

Tests were then run with batch sizes of 16 and 32 using the "LightSimbackend". The batch size 16 configuration followed the same path as the "Do Nothing" agent and ultimately failed to survive due to its inaction. In contrast, the batch size 32 configuration successfully survived by immediately reconnecting line 17 at time step 1, which was the

only action taken. This promptness made this configuration the one with the highest reward of 11948.

The last test performed was to change the network architecture from three hidden layers with 200 neurons to two hidden layers with the same neurons, remaining with the backend "LightSimBackend" and the baseline hyperparameters. The agent failed to complete the task of surviving all time steps, remaining inactive and functioning only until time step 113.

Tables 11 through 13 summarize the results for this scenario, in which the grid begins with a critical power line disconnected.

TABLE 11. Do Nothing Agents Results

Do Nothing

Backend	Survival	Reward	Distinct Actions
PandaPowerBackend	113	2070	0
LightSimBackend	113	2070	0

TABLE 12. PPO Agents Results

PPO

	Backend	Policy	Learning Rate	Batch Size	Policy_kwargs	Survival	Reward	Distinct Actions	Reconnect Line 17?	Line Reconnected at Time Step:
	LightSimBackend	MlpPolicy	1.00E-03	16	[200, 200, 200]	113	2070	1	No	
	LightSimBackend	MlpLstmPolicy	1.00E-03	16	[200, 200, 200]	113	2070	0	No	
	LightSimBackend	MlpPolicy	1.00E-03	32	[200, 200, 200]	113	2070	0	No	
	LightSimBackend	MlpLstmPolicy	1.00E-03	32	[200, 200, 200]	113	2070	0	No	
*	LightSimBackend	MlpLstmPolicy	1.00E-03	16	[200, 200, 200]	Yes	11887	2	Yes	1

* n_epochs = 500

TABLE 13. SAC Agents Results

SAC

Backend	Policy	Learning Rate	Batch Size	Policy_kwargs	Survival	Reward	Distinct Actions	Reconnect Line 17?	Line Reconnected at Time Step:
PandaPowerBackend	MlpPolicy	1.00E-03	8	[200, 200, 200]	Yes	11942	1	Yes	35
LightSimBackend	MlpPolicy	1.00E-03	8	[200, 200, 200]	Yes	11947	1	Yes	3
LightSimBackend	MlpPolicy	1.00E-04	8	[200, 200, 200]	113	2070	0	No	
LightSimBackend	MlpPolicy	1.00E-03	16	[200, 200, 200]	113	2070	0	No	
LightSimBackend	MlpPolicy	1.00E-03	32	[200, 200, 200]	Yes	11948	1	Yes	1
LightSimBackend	MlpPolicy	1.00E-03	8	[200, 200]	113	2070	0	No	

CHAPTER 6

Conclusion and Future Work

6.1. Conclusion

The aim of this work was to explore the effectiveness of deep reinforcement learning techniques for autonomous topology control in the Portuguese power grid system. Our main objective was to optimise grid stability and performance while minimising manual intervention, using different agent configurations. Through systematic testing, it was sought to compare these agents, PPO and SAC, against a baseline "Do Nothing" agent, using a set of well-defined KPIs such as survival rate, rewards, topology changes and the number of reconnections and disconnections in the grid.

The "Do Nothing" agent served as our baseline, and the results showed that when tested on a reference grid without any constraints, the grid remained stable without any intervention, achieving a reward of 11948. These results confirm the inherent stability of the network under the conditions tested. To further evaluate the resilience of the system, a second scenario was implemented that forced the disconnection of a critical line. In this scenario, the agent was unable to maintain stability, surviving only 113 time steps before exceeding the thermal limits of the power lines. The backend used, whether "PandaPowerBackend" or "LightSimBackend", had no noticeable effect on performance.

The PPO agent performed poorly in the configurations tested. In the baseline Scenario 1, even the "Do Nothing" agent survived, but about half of these agents configurations made inappropriate decisions by requesting to disconnect power lines, leading to the failure. The four configurations that remained stable in the baseline scenario used LightSimBackend, a learning rate of 1E-3, three hidden layers of 200 neurons and the combination of a batch size of 16 and 32 and the policies "MlpPolicy" and "MlpLstmPolicy". When applied to the scenario with an initial line disconnection, these agents maintained its strategy of inaction and not reconnecting the disconnected line, which ultimately led to the power line capacity being exceeded. This behaviour suggests that PPO agents may have found it difficult to learn effective strategies for controlling the power grid, especially in scenarios with significant disturbances or unexpected events. With this in mind, the number of epochs was increased to 500 to improve model convergence, and with this action the agent was able to reconnect the disconnected line and complete the test steps without exceeding the line capacity. Although he has achieved this result, with a large number of training epochs there is a risk of overfitting to the specific batch of data and this may result in poor generalisation to new data from the environment.

SAC agents demonstrated significantly greater adaptability to network conditions compared to PPO agents. In the first scenario, all SAC agent configurations successfully maintained stability throughout the entire testing process without implementing any topological changes. This demonstrates the SAC agent's ability to maintain the current grid state without intervention. Although these baseline results are promising, they do not reveal the full potential of the SAC agent. It was essential to evaluate the agent's ability to adapt dynamically to evolving network conditions and take proactive measures to optimise performance. In the new scenario with an initial line disconnection, only three configurations successfully survived all time steps. Two of these configurations used baseline hyperparameters and simply switched backends. "LightSimBackend" emerged as the preferred backend, demonstrating slightly more effective performance than "PandaPowerBackend" and achieving the optimal grid topology more rapidly at time step 3. The other configuration, which used a batch size of 32 instead of 8, successfully reconnected the power line at the very first time step. These results highlight the superior adaptability of the SAC agent and its potential for effective power grid control, even in challenging scenarios with initial line failures.

In conclusion, and in response to the research questions posed, this study demonstrates the potential of Deep Reinforcement Learning (DRL) to autonomously manage a simplified model of the Portuguese electricity grid and optimize ATC but further investigation is necessary to draw definitive conclusions (Q1). The results provide valuable insights into the viability of AI-driven topology control for real-time network optimisation. However, the limitations of our simplified model and the complexity of the real Portuguese electricity network suggest that further research is needed to fully evaluate the applicability of DRL in a real-world setting.

In terms of the performance of different DRL agents (Q2), SAC emerged as the most effective, exceeding PPO in terms of adaptability, stability and efficiency. This superior performance could be attributed to SAC's off-policy learning capability, which allows it to learn not only from current interactions but also from past experience. This feature enables SAC to reuse old data, making it more efficient than on-policy methods such as PPO, which rely solely on real-time data and may be less adaptable to the dynamic needs of the grid.

For those who wish to reproduce or extend the results, the code used for this dissertation is available at:
https://github.com/diogoemlopes/Dissertation_DiogoLopes_2024

6.2. Future Work

The results of this study highlight several promising directions for future research. A key area for improvement is gathering more data to improve the training of the models, helping the agents to make more informed decisions, and extending the testing period to assess longer-term performance. Additionally, fine-tuning the deep reinforcement learning agents in that new context through further experimentation with hyperparameters could

help reduce unnecessary topological changes while maintaining grid stability. This fine-tuning could be particularly important when applying the agents to more diverse scenarios, where adjustments in hyperparameters may be necessary to ensure optimal performance across varying conditions.

Furthermore, future work should also consider the costs associated with every action taken by the agents, ensuring that optimisation efforts balance performance and operational efficiency. It is important to account for the costs related to adjusting energy generation to match supply with demand.

Real-time deployment remains the ultimate goal and testing these agents on more complex/real-world power grid models, such as a more detailed Portuguese network, is crucial for practical application.

References

Alhazmi, M., Dehghanian, P., Wang, S., & Shinde, B. (2019). Power Grid Optimal Topology Control Considering Correlations of System Uncertainties. *IEEE Transactions on Industry Applications*, 55(6), 5594–5604. https://doi.org/10.1109/TIA.2019.2934706

Alshamrani, A. M., El-Meligy, M. A., Sharaf, M. A. F., Mohammed Saif, W. A., & Awwad, E. M. (2023). Transmission Expansion Planning Considering a High Share of Wind Power to Maximize Available Transfer Capability. *IEEE Access*, 11, 23136–23145. https://doi.org/10.1109/ACCESS.2023.3253201

Appelrath, H.-J., Terzidis, O., & Weinhardt, C. (2012). Internet of Energy. *Business & Information Systems Engineering*, 4(1), 1–2. https://doi.org/10.1007/s12599-011-0197-x

Chauhan, R., Naresh, R., Kenedy, M., & Aharwar, A. (2023). A streamlined and enhanced iterative method for analysing power system available transfer capability and security. *Electric Power Systems Research*, 223, 109528. https://doi.org/10.1016/j.epsr.2023.109528

Dehghanian, P., Wang, Y., Gurrala, G., Moreno-Centeno, E., & Kezunovic, M. (2015). Flexible implementation of power system corrective topology control. *Electric Power Systems Research*, 128, 79–89. https://doi.org/10.1016/j.epsr.2015.07.001

Dillon, J. V., Langmore, I., Tran, D., Brevdo, E., Vasudevan, S., Moore, D., Patton, B., Alemi, A., Hoffman, M., & Saurous, R. A. (2017). TensorFlow Distributions.

Engstrom, L., Ilyas, A., Santurkar, S., Tsipras, D., Janoos, F., Rudolph, L., & Madry, A. (2019). Implementation Matters in Deep RL: A Case Study on PPO and TRPO.

Fakoor, R., Chaudhari, P., & Smola, A. J. (2020). P3O: Policy-on Policy-off Policy Optimization. https://github.com/rasoolfa/P3O.

Fang, X., Misra, S., Xue, G., & Yang, D. (2012). Smart Grid — The New and Improved Power Grid: A Survey. *IEEE Communications Surveys & Tutorials*, 14(4), 944–980. https://doi.org/10.1109/SURV.2011.101911.00087

Fisher, E. B., O'Neill, R. P., & Ferris, M. C. (2008). SmartOptimal Transmission Switching. *IEEE Transactions on Power Systems*, 23(3), 1346–1355. https://doi.org/10.1109/TPWRS.2008.922256

Fuller, J. D., Ramasra, R., & Cha, A. (2012). Fast Heuristics for Transmission-Line Switching. *IEEE Transactions on Power Systems*, 27(3), 1377–1386. https://doi.org/10.1109/TPWRS.2012.2186155

Glavitsch, H. (1985). Switching as means of control in the power system. *International Journal of Electrical Power & Energy Systems*, 7(2), 92–100. https://doi.org/10.1016/0142-0615(85)90014-6

González Vázquez, J. M., Sauer, J., & Appelrath, H.-J. (2012). Methods to Manage Information Sources for Software Product Managers in the Energy Market. *Business & Information Systems Engineering*, 4(1), 3–14. https://doi.org/10.1007/s12599-011-0200-6

Gravener, M. H., & Nwankpa, C. (1999). Available transfer capability and first order sensitivity. *IEEE Transactions on Power Systems*, 14(2), 512–518. https://doi.org/10.1109/59.761874

Haarnoja, T., Zhou, A., Hartikainen, K., Tucker, G., Ha, S., Tan, J., Kumar, V., Zhu, H., Gupta, A., Abbeel, P., & Levine, S. (2018). Soft Actor-Critic Algorithms and Applications.

Henderson, M. I., Novosel, D., & Crow, M. L. (2017). Electric Power Grid Modernization Trends, Challenges, and Opportunities. *IEEE*.

Huang, Q., Huang, R., Hao, W., Tan, J., Fan, R., & Huang, Z. (2020). Adaptive Power System Emergency Control Using Deep Reinforcement Learning. *IEEE Transactions on Smart Grid*, 11(2), 1171–1182. https://doi.org/10.1109/TSG.2019.2933191

Kadiyala, A., & Kumar, A. (2017). Applications of Python to evaluate environmental data science problems. *Environmental Progress & Sustainable Energy*, 36(6), 1580–1586. https://doi.org/10.1002/ep.12786

Kaluarachchi, T., Reis, A., & Nanayakkara, S. (2021). A review of recent deep learning approaches in human-centered machine learning. In *Sensors* (Vol. 21, Issue 7). MDPI AG. https://doi.org/10.3390/s21072514

Kelly, A., O'Sullivan, A., de Mars, P., & Marot, A. (2020). Reinforcement Learning for Electricity Network Operation. *ArXiv*. http://arxiv.org/abs/2003.07339

Kitchenham. (2007). Guidelines for performing Systematic Literature Reviews in *Software Engineering*.

Lan, T., Duan, J., Zhang, B., Shi, D., Wang, Z., Diao, R., & Zhang, X. (2020). AI-Based Autonomous Line Flow Control via Topology Adjustment for Maximizing Time-Series ATCs. *IEEE PES GM 2020 (Preprint)*.

Marot, A., Donnot, B., Romero, C., Veyrin-Forrer, L., Lerousseau, M., Donon, B., & Guyon, I. (2020). Learning to run a power network challenge for training topology controllers. http://arxiv.org/abs/1912.04211

Mazi, A. A., Wollenberg, B. F., & Hesse, M. H. (1986). Corrective Control of Power System Flows by Line and Bus-Bar Switching. *IEEE Transactions on Power Systems*, 1(3), 258–264. https://doi.org/10.1109/TPWRS.1986.4334990

Nardelli, P. H. J., Rubido, N., Wang, C., Baptista, M. S., Pomalaza-Raez, C., Cardieri, P., & Latva-aho, M. (2014). Models for the modern power grid. *The European Physical Journal Special Topics*, 223(12), 2423–2437. https://doi.org/10.1140/epjst/e2014-02219-6

Ottmar Edenhofer, Ramón Pichs Madruga, Youba Sokona, Kristin Seyboth, Patrick Matschoss, Susanne Kadner, Timm Zwickel, Patrick Eickemeier, Gerrit Hansen, Steffen Schlomer, & Christoph von Stechow. (2012). Renewable energy sources and climate change mitigation: Special report of the intergovernmental panel on climate change. *Cambridge University Press*.

Panda, S., Mohanty, S., Rout, P. K., Sahu, B. K., Parida, S. M., Samanta, I. S., Bajaj, M., Piecha, M., Blazek, V., & Prokop, L. (2023). A comprehensive review on demand side management and market design for renewable energy support and integration. In *Energy Reports* (Vol. 10, pp. 2228–2250). Elsevier Ltd. https://doi.org/10.1016/j.egyr.2023.09.049

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., Facebook, Z. D., Research, A. I., Lin, Z., Desmaison, A., Antiga, L., Srl, O., & Lerer, A. (2017). Automatic differentiation in PyTorch.

Rahman, A., Farrok, O., & Haque, M. M. (2022). Environmental impact of renewable energy source based electrical power plants: Solar, wind, hydroelectric, biomass, geothermal, tidal, ocean, and osmotic. *Renewable and Sustainable Energy Reviews*, 161, 112279. https://doi.org/10.1016/j.rser.2022.112279

RTE France. (2024). Grid2Op. Https://Github.Com/Rte-France/Grid2Op/Blob/Master/README.Md.

Saabith, A. L. S., Fareez, M., & Vinothraj, T. (2019). PYTHON CURRENT TREND APPLICATIONS-AN OVERVIEW POPULAR WEB DEVELOPMENT FRAMEWORKS IN PYTHON. *International Journal of Advance Engineering and Research Development*, 6(10).

Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal Policy Optimization Algorithms.

Venkat, A. N., Hiskens, I. A., Rawlings, J. B., & Wright, S. J. (2008). Distributed MPC Strategies With Application to Power System Automatic Generation Control. *IEEE Transactions on Control Systems Technology*, 16(6), 1192–1206. https://doi.org/10.1109/TCST.2008.919414

Xiao, H., & Cao, M. (2020). Balancing the demand and supply of a power grid system via reliability modeling and maintenance optimization. *Energy*, 210, 118470. https://doi.org/10.1016/j.energy.2020.118470

Yoon, D., Hong, S., Lee, B.-J., & Kim, K.-E. (2021). WINNING THE L2RPN CHALLENGE: POWER GRID MANAGEMENT VIA SEMI-MARKOV AFTERSTATE ACTOR-CRITIC.

Zhao, C., Topcu, U., Li, N., & Low, S. (2014). Design and Stability of Load-Side Primary Frequency Control in Power Systems. *IEEE Transactions on Automatic Control*, 59(5), 1177–1189. https://doi.org/10.1109/TAC.2014.2298140