IMOS
Semester Project

Power Grid Vulnerability Analysis via Reinforcement Learning

Author:
BAHUREL Benjamin 326888

Supervisors:
Prof. Olga Fink, Ismail Nejjar

Robotics Section - MA2

Contents

1 Introduction

This project was built on the thesis "Power System Security Analysis using Reinforcement Learning" by José María Sunyer Nestares, which explored the use of reinforcement learning to identify vulnerable lines in power transmission networks. The central idea was to simulate the sequential disconnection of lines and observe how these actions could trigger cascading failures across the grid and possibly its complete blackout.

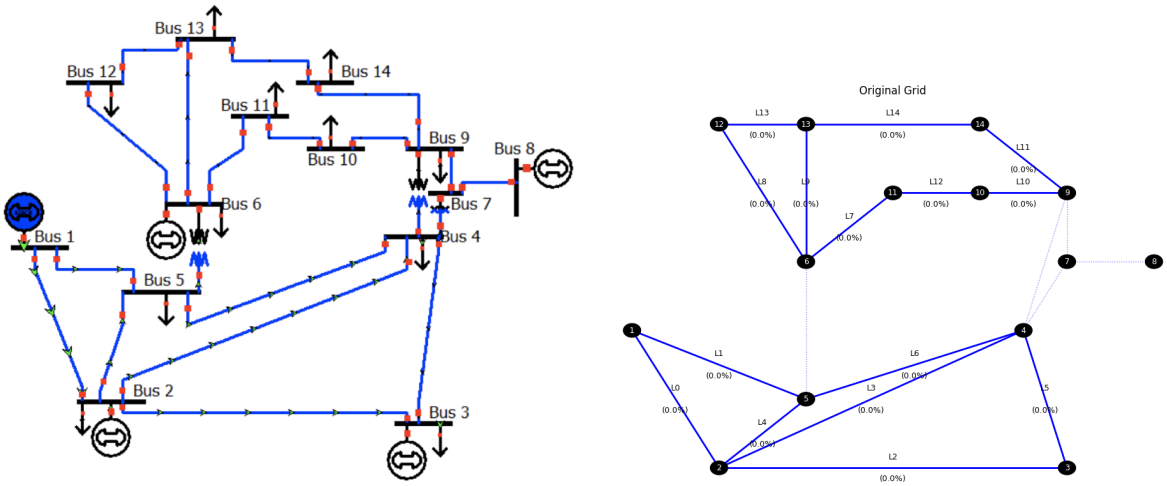
As part of my Master's studies in Robotics at EPFL, I implemented a reproduction of the thesis framework using the IEEE 14-bus system and the Pandapower library. The agent logic and environment structure were kept close to the original, with a Monte Carlo algorithm guiding the action-value updates. However, a key modification was made to the reward function. In the original thesis, the grid failure was observed but not explicitly used in the reward signal. In this version, I introduced a reward that favors actions leading to rapid system collapse, with the idea of training the agent to identify highly critical disconnection sequences more efficiently. This change provides a different optimization signal and strengthens the agent's ability to uncover minimal sets of lines that can compromise grid stability.

The report details the reproduction process, the core idea behind this adaptation, and the implications for power system vulnerability analysis.

2 Contingency Analysis

To better understand the structure and behavior of the IEEE 14-bus system, I began by conducting a contingency analysis for N-1, N-2, and N-3 scenarios. This involved systematically disconnecting one, two, or three transmission lines and observing the grid's ability to maintain a stable power flow. The goal was to observe and analyze which lines play a critical role in network stability and at which point grid failure was happening. This preliminary step was useful not only for validating the simulation setup but also for interpreting the agent's behavior later in the learning phase.

Based on the contingency analysis, the focus will be limited to disconnecting standard transmission lines, excluding transformer disconnections from the set of possible actions. Below are the visual representations of the IEEE 14-bus system that will be used to carry out this analysis :



(a) ICSEG Representation

(b) Pandapower Representation

Figure 1: Representations of the IEEE 14-Bus system's grid

2.1 N-1 Analysis

The N-1 analysis consists in evaluating the grid's response to the failure of a single transmission line at a time. For each line, a power flow calculation is performed after its disconnection to determine whether the system can still operate normally or if it results in a failure.

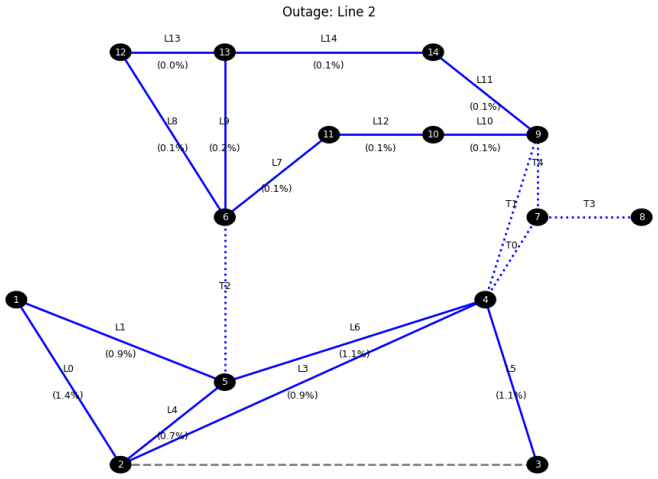


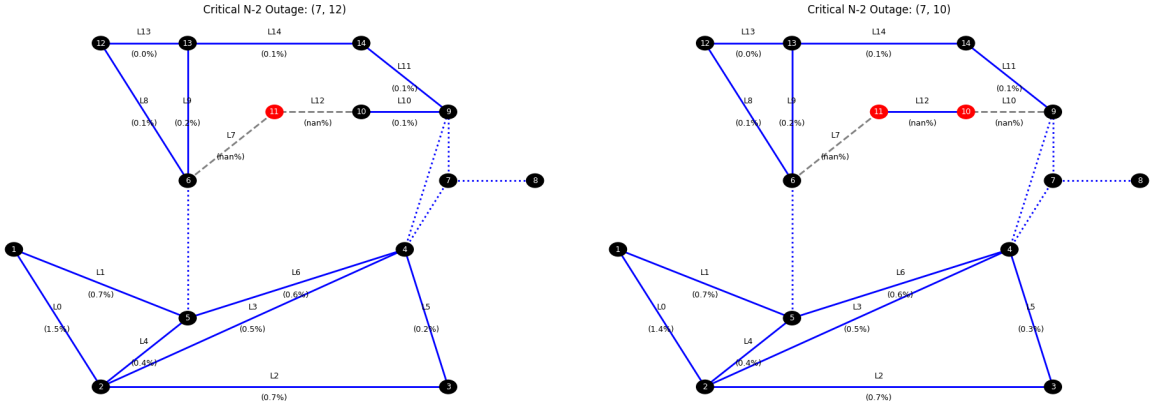
Figure 2: Example of simulation with line 2 disconnected (dashed gray)

The results of the N-1 contingency analysis show that the IEEE 14-bus system remains stable regardless of which individual transmission line is disconnected. This robustness can be attributed to several structural properties of the grid. First, the network is highly meshed, offering multiple alternative paths for power to flow in case of a local failure. Second, no single line acts as a critical bottleneck, which helps to prevent islanding or overloading elsewhere in the system.

More generally, this behavior reflects a fundamental design principle in power system engineering: real-world transmission networks are generally built to withstand any single failure without compromising stability. N-1 security is a standard reliability criterion that guides the design and operation of most power grids.

2.2 N-2 Analysis

The N-2 analysis extends this approach by evaluating the grid's response to the simultaneous disconnection of two transmission lines. For each pair of lines, a power flow simulation is run to assess whether the system can maintain a stable operating point or if the double failure leads to the grid's blackout.



(a) Islanding of 1 bus (bus 11 in red)

(b) Islanding of 2 buses (11 and 10 in red)

Figure 3: Examples of islanding due to double line disconnection

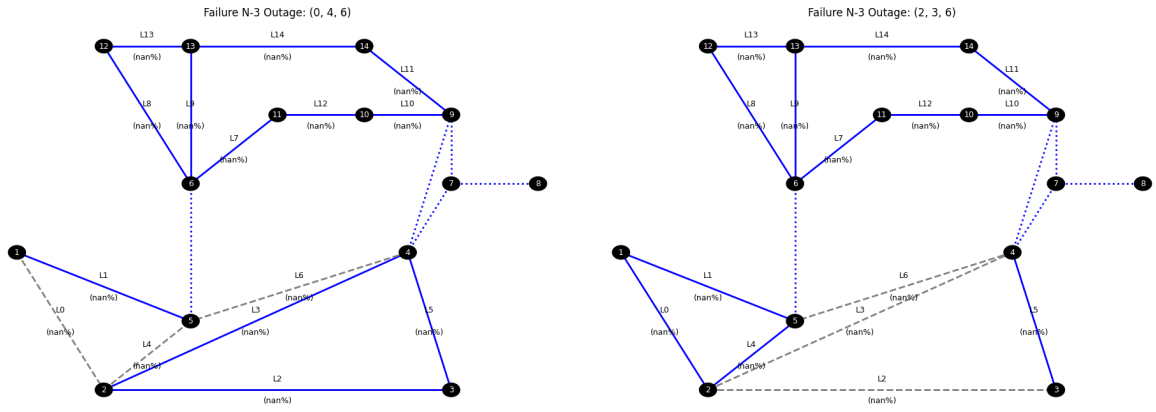
The N-2 contingency analysis reveals the emergence of a new type of issue: islanding. In a few specific line pair disconnections, one or more buses become electrically isolated from the rest of the network.

Out of the 105 possible combinations, 9 cases (approximately 8.6%) resulted in such critical configurations. For the purposes of this study, islanding is not considered a failure in the strict sense (the simulation is still technically feasible) but rather a critical structural problem that indicates a weakened connectivity in the grid.

In real-world scenarios, the impact of islanding would depend heavily on the type of load and the local generation capabilities. However, in the context of this analysis, it is treated as a structural vulnerability worth highlighting, even if it does not lead to a full system collapse.

2.3 N-3 Analysis

The N-3 analysis explores the grid's response to more extreme conditions by disconnecting three transmission lines at once. The goal is to assess whether this level of simultaneous disconnection is enough to trigger a grid failure or cause significant cascading effects. This step helps evaluate how much stress the network can tolerate before losing stability.



(a) Combination (0,4,6)

(b) Combination (2,3,6)

Figure 4: Examples of combination that leads to grid's failure

The N-3 analysis reveals the first clear signs of grid failure. Among the 455 possible combinations of three simultaneous line disconnections, two specific cases (0, 4, 6) and (2, 3, 6) cause the power flow computation to fail entirely.

While these failure cases remain rare (0.44%), the number of critical cases increases significantly: 114 combinations (25.05%) result in islanding or voltage and line overload violation issues.

This marks a turning point in the analysis; from three simultaneous disconnections onwards, the grid begins to show real signs of vulnerability. These findings suggest that while the IEEE 14-bus system is resilient under standard contingency levels, its stability quickly degrades when exposed to higher-order failures, highlighting the importance of identifying such high-risk combinations in advance.

With this understanding of how the grid responds to increasing levels of stress, we now turn to the core of the project: reproducing the reinforcement learning framework proposed in the thesis.

3 RL Implementation

The objective of the reinforcement learning implementation is to train an agent capable of identifying the most critical combinations of line disconnections that can lead to cascading failures and eventually cause the grid to collapse.

The agent is designed to act as an external attacker, interacting with the environment, here modeled as the power grid, by sequentially disconnecting transmission lines. Each action modifies the grid's state and can lead to further automatic line outages caused by overloads, effectively simulating the behavior of real-world cascading failures. Over multiple episodes, the agent learns to identify the most efficient sequences of disconnections that drive the system toward instability.

Since this work is based on José María Sunyer Nestares' thesis "Power System Security Analysis using Reinforcement Learning", the reinforcement learning framework closely follows the original implementation. Most of the agent-environment structure and learning logic are directly inspired by the thesis. However, one key difference introduced in this project lies in the definition of the reward function, which has been adapted to better guide the agent toward discovering minimal disconnection sequences that can trigger a full grid collapse. This modification and its impact will be discussed in more detail in the following section.

3.1 Reward Functions

3.1.1 Thesis

The reward function, as defined in the thesis, quantifies the evolution of cascading events. It is based on the change in the number of disconnected lines between two consecutive time steps and is expressed as:

$$r_t = |\Delta \mathcal{C}_{t+1}| \tag{1}$$

This formulation reflects the agent's objective to trigger large cascading failures. The higher the value of r_t, the more impactful the action taken at time t.

Accordingly, the return function is defined as the sum of discounted future rewards, starting from time t:

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots = \sum_{k=0}^{T} \gamma^k R_{t+1+k} \tag{2}$$

which also satisfies the recursive relation:

$$G_t = R_{t+1} + \gamma G_{t+1} \tag{3}$$

In summary, in the thesis, the agent will search for the actions which maximize the total number of cascaded lines over time, in alignment with the reward definition above. Even though a cascade may involve a large number of line disconnections, it does not always result in a full collapse of the grid.

3.1.2 Implementation

In this implementation, the reward function has been redefined to encourage the agent to actively search for disconnection sequences that lead to rapid and large-scale grid failure. Unlike the original approach, which rewards every incremental cascade event, the modified reward focuses on the number of lines that are automatically disconnected after the agent's action, due to overloads. The goal is to reward actions that trigger immediate cascading effects.

Formally, at each timestep t, after applying an action a_t, the reward r_t is defined as:

$$r_t = |\Delta \mathcal{C}_{t+1}| \tag{4}$$

where $\mathcal{C}_{t+1}$ is the set of new lines that go out of service due to overloads during the cascading process triggered by a_t. The longer and more destructive the cascade, the higher the reward.

In the event of a complete grid collapse (i.e., when the power flow solver fails to converge), the agent receives a reward equivalent to the number of active lines before the grid failure :

$$r_{t_{failure}} = N_{\text{active}} \quad \text{(in case of total failure)} \tag{5}$$

This formulation encourages the agent to find minimal sequences of disconnections that lead to maximal disruption, making the learning process more focused on discovering critical vulnerabilities rather than simply accumulating small cascades over time.

So then the return function for a sequence that leads to grid failure will be :

$$G_t = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots + \gamma^n R_{t_{failure}} = R_{t+1} + \gamma R_{t+2} + \gamma^2 R_{t+3} + \cdots + \gamma^n N_{active} \tag{6}$$

This means that the longer the agent takes to trigger a collapse, the more the final reward is discounted, which naturally encourages the discovery of shorter and more efficient sequences leading to grid failure.

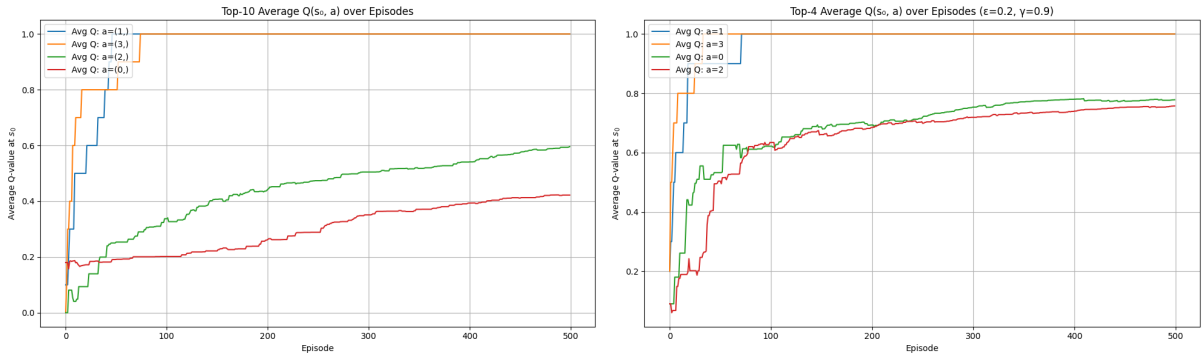
3.2 Results

To evaluate the performance of both the original reinforcement learning framework from the thesis and the modified version developed in this project, I tested the two approaches on three different power grid models: `case14`, `case4gs`, and `case39`. The latter two were already part of the original thesis, while `case14` was specifically chosen for this project in order to gain a deeper understanding of a grid that had not been previously explored.

For each grid, we analyze how effectively the agent is able to identify high-risk disconnection sequences and compare the results between the baseline and modified reward formulation. The following sections present the outcomes for each grid individually.

3.2.1 Case4gs

The 4-bus system serves as a reference case to verify the correctness of the reproduced reinforcement learning framework. As shown in the figure, the Q-value curves from my implementation closely follow those obtained in the original thesis. The same actions, particularly disconnections of lines 1 and 3, are consistently identified as the most impactful early in training. While the overall simulation time is slightly longer in my version (approximately 30 seconds per trial compared to 20 seconds in the thesis), the learning dynamics and final action rankings are in clear agreement.



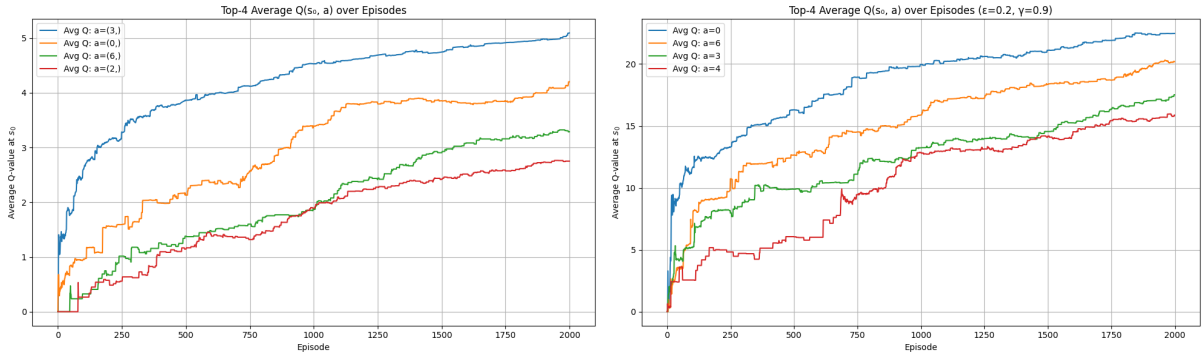
(a) Result of thesis implementation for `case4gs` grid (b) Result of own implementation for `case4gs` grid

Figure 5: Implementations results for `case4gs`

This validates that the structure of the environment and the learning behavior of the agent have been correctly re-implemented and are functioning as intended.

3.2.2 Case14

The 14-bus system was used to further evaluate the agent's ability to generalize to a more complex and redundant network.



(a) Result of thesis implementation for `case14` grid

(b) Result of own implementation for `case14` grid

Figure 6: Implementations results for `case14`

Thesis Q-values

Action: a	Q
3	5.092
0	4.202
6	3.288
2	2.751
4	2.652
9	2.527
10	2.493
7	2.333
8	1.845
5	1.748
12	1.705
13	1.508
11	1.036
1	0.619
14	0.428

My Implementation Q-values

Action: a	Q
0	22.464
6	20.189
3	17.516
4	15.870
2	14.158
5	12.921
9	12.609
13	11.225
12	8.683
8	8.219
10	8.069
7	7.671
1	5.509
11	5.246
14	4.933

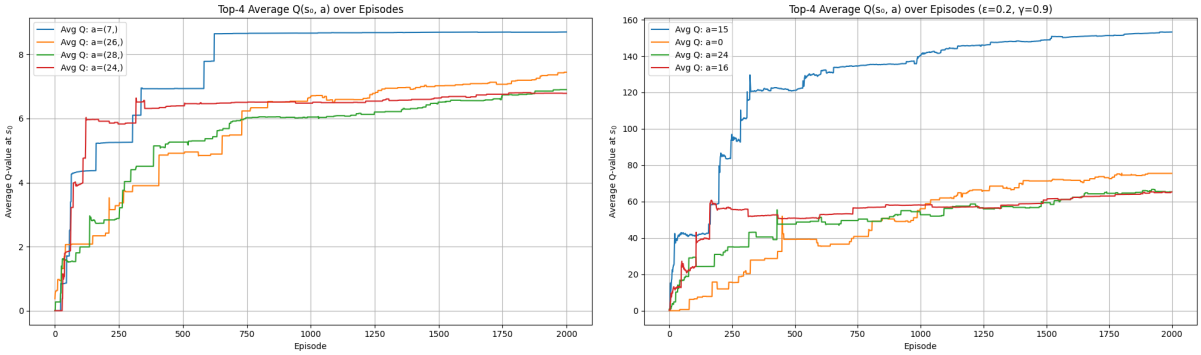
Table 1: Top Q-values at s_0 for `case14`: comparison between the thesis implementation and my own

Both the thesis implementation and my own version converge toward identifying the same set of high-impact actions, with lines 0, 2, 3, 4, and 6 consistently ranked among the top. These results are fully aligned with the N-3 contingency analysis conducted earlier, which had already highlighted these lines as structurally critical.

Despite slightly longer runtime in my implementation, the final policy behavior matches the reference, which reinforces the validity of the reproduced framework.

3.2.3 Case39

The 39-bus system was used as a final benchmark to evaluate the scalability of the reinforcement learning framework on a larger and more complex grid. In the thesis implementation, the agent consistently identified actions on lines 7, 26, 28, and 24 as the most impactful. In contrast, my implementation converged toward a different set of top actions, with line 15 emerging as the dominant choice, followed by lines 0, 24, and 16. While some overlap remains, such as the presence of line 24 in both runs, the overall ranking of actions differs.



(a) Result of thesis implementation for `case39` grid

(b) Result of own implementation for `case39` grid

Figure 7: Implementations results for `case39`

In the case of the 39-bus system, the set of critical lines identified by my implementation diverges from those found in the original thesis, with the exception of a single common line. This difference can be explained by several factors. First, the reward structure in my version explicitly includes a bonus when the grid collapses, which pushes the agent to prioritize fast and efficient destabilization. In contrast, the thesis formulation rewards the overall number of cascading failures without specifically targeting complete system failure. This change alters the learning objective and can lead the agent to favor different disconnection strategies. Additionally, the 39-bus grid introduces much more complexity and variability than the smaller cases, which increases the sensitivity of the learning process to early action choices and exploration paths.

As a result, two agents trained under slightly different conditions can converge toward different, but still valid, high-risk line combinations.

3.3 Limitations

While the reproduced framework provides meaningful insights into power grid vulnerabilities, several limitations remain.

First, the current setup is entirely grid-specific : the agent does not learn a reusable policy or model, and all learning is done from scratch for each new grid. This makes it difficult to generalize or transfer knowledge across network topologies.

Second, the Monte Carlo algorithm used here relies on episodic sampling, which means it only explores a fraction of the state-action space. If the number of episodes is limited, the agent may converge to suboptimal solutions simply due to insufficient coverage. Moreover, there is no guarantee that the policy will converge, particularly in large and complex environments like the 39-bus system, a challenge that was already acknowledged in the original thesis.

4 Conclusion

This project was set out to reproduce and extend the reinforcement learning framework introduced in José María Sunyer Nestares' thesis, with the goal of identifying critical vulnerabilities in power transmission networks. I began by performing a detailed contingency analysis of the IEEE 14-bus system to understand how structural weaknesses appear when the grid is increasingly stressed. This initial step justified the design and interpretation of the reinforcement learning experiments that followed.

The RL framework was reconstructed using a Monte Carlo approach, with a key modification introduced in the reward function to better target rapid and large-scale grid failures. By comparing the outcomes across three different network topologies : `case4gs`, `case14`, and `case39` ; I observed strong agreement with the thesis results in smaller grids, and noted meaningful divergences in the larger `case39` system. These differences reflect the increased sensitivity and complexity that arise when scaling reinforcement learning to more realistic environments, as well as the influence of reward shaping on agent behavior.

Despite certain limitations, such as the grid-specific nature of the current setup and the sampling inefficiencies of Monte Carlo learning, this work demonstrates the feasibility of using reinforcement learning to uncover hidden structural weaknesses in power grids. Looking ahead, a promising direction would be to move towards real reinforcement learning by implementing an agent that learns a generalizable policy, which could then be deployed on any grid configuration. This project lays a solid foundation for such future developments.

5 References

[1] Sunyer Nestares, José María. *Power System Security Analysis using Reinforcement Learning*. University of Illinois at Urbana-Champaign, May 2020.

[2] IEEE 4-Bus system grid : *https://icseg.iti.illinois.edu/ieee-14-bus-system/*

[3] IEEE 14-Bus system grid : *https://matpower.org/docs/ref/matpower5.0/case4gs.html*

[4] IEEE 39-Bus system grid : *https://icseg.iti.illinois.edu/ieee-39-bus-system/*

[5] PYPOWER Format : *https://rwl.github.io/PYPOWER/api/pypower.caseformat-module.html*

[6] Pandapower Documentation : *https://www.pandapower.org/about/#analysis*

[7] Markov Decision Process Documentation : *https://medium.com/@alex_pimenov/rl-part-3-markov-decision-process-policy-bellman-optimality-equation-8b83ee670037*