Learning to run a Power Network Challenge: a Retrospective Analysis

Antoine Marot *AI Lab, Reseau de Transport d'Electricite (RTE) France*
Benjamin Donnot *AI Lab, Reseau de Transport d'Electricite (RTE) France*
Gabriel Dulac-Arnold *Google Research,Brain Team, Google France*
Adrian Kelly *Electric Power Research Institute (EPRI) Ireland*
Aidan O'Sullivan *UCL Energy Institute, University College of London*
Jan Viebahn *Digital and Process Excellence, TenneT*
Mariette Awad *Electrical & Computer Engineering Department, American University of Beirut*
Isabelle Guyon *Chalearn & University Paris-Saclay France*
Patrick Panciatici *R&D Department, Reseau de Transport d'Electricite (RTE) France*
Camilo Romero *CELEC EP Transelectric Ecuador*

Editor: Hugo Jair Escalante and Katja Hofmann

Abstract

Power networks, responsible for transporting electricity across large geographical regions, are complex infrastructures on which modern life critically depend. Variations in demand and production profiles, with increasing renewable energy integration, as well as the high voltage network technology, constitute a real challenge for human operators when optimizing electricity transportation while avoiding blackouts. Motivated to investigate the potential of Artificial Intelligence methods in enabling adaptability in power network operation, we have designed a L2RPN challenge to encourage the development of reinforcement learning solutions to key problems present in the next-generation power networks. The NeurIPS 2020 competition was well received by the international community attracting over 300 participants worldwide.

The main contribution of this challenge is our proposed comprehensive 'Grid2Op' framework, and associated benchmark, which plays realistic sequential network operations scenarios. The Grid2Op framework, which is open-source and easily re-usable, allows users to define new environments with its companion GridAlive ecosystem. Grid2Op relies on existing non-linear physical power network simulators and let users create a series of perturbations and challenges that are representative of two important problems: a) the uncertainty resulting from the increased use of unpredictable renewable energy sources, and b) the robustness required with contingent line disconnections.

In this paper, we give the highlights of the NeurIPS 2020 competition. We present the benchmark suite and analyse the winning solutions, including one super-human performance demonstration. We propose our organizational insights for a successful competition and conclude on open research avenues. Given the challenge success, we expect our work will foster research to create more sustainable solutions for power network operations.

Keywords: Power Networks, Reinforcement Learning, Decision-making, Competition

1. Introduction

To meet climate goals, power networks are relying more and more on renewable energy generation and decentralising their production. The traditional tools used by electrical

engineers to solve network issues are becoming increasingly inadequate. For further reading on the challenges of future power network operations, readers can refer to the accompanying white paper which introduces the problem, potential solutions (Kelly et al., 2020).

There are some promising innovations and recent applications of machine learning (ML) techniques that may offer solutions to these challenges. However the opportunities for ML scientists and researchers to work on the problem was limited by a lack of usable environments, baselines, data, networks and simulators.

The "Learning to run a power network" (L2RPN) challenge is a series of competitions that model the sequential decision-making environments of real-time power network operations as illustrated in Figure 1. It is backed by the open-source GridAlive[1] ecosystem and Grid2op[2] core framework.

The aim of L2RPN is to develop a catalyst to foster faster progress in the field, building on advances in AI and reinforcement learning (RL) such as the ImageNet benchmark (Deng et al., 2009) for computer vision, and many others in the last decade. The L2RPN challenge aimed at making new large benchmarks for solutions to the real-world problem of complex network operations readily available through the unique Grid2Op framework. The Grid2Op framework has raised awareness of the challenge beyond the power system community, to new, diverse and young communities of scientists and researchers. Additionally, the challenges present an opportunity for the AI community to demonstrate recent breakthroughs which could successfully find applications in other real-world problems, and to eventually go beyond well-studied game environments (Brown and Sandholm, 2017; Silver et al., 2017; Vinyals et al., 2019; Badia et al., 2020). Conversely, the aim is to raise awareness among the power system community about the innovations and potentials of AI and RL algorithms to solve network challenges and to embrace the application of different approaches to traditional problems (Prostejovsky et al., 2019).

Although RL approaches have successfully integrated some real-world applications recently in robot walk (Li et al., 2021), autonomous vehicle (Cohen, 2020) or stratospheric balloon navigation, it has only remained as a research niche in power-systems (Glavic et al., 2017) through years. This is probably because of the lack of both a substantially large benchmark and a widely used framework that we address through L2RPN. Advocates for real-world RL (Kidziński et al., 2020; Dulac-Arnold et al., 2021; Mohanty et al., 2020) have indeed made advances in recent years by developing related challenges such as in biomechanics and train logistics. L2RPN also pushes boundaries by framing the challenges over large spatio-temporal decision-making spaces. Foreseen industrial application could further come as a smart operation recommender system for operators, similar to the one developed for Google data center cooling (Jim Gao, 2018).

While initial competitions in L2RPN tested the feasibility of developing realistic power network environments (Marot et al., 2020a) and the applicability of RL agents (Lan et al., 2020; Yoon et al., 2021; Subramanian et al., 2020), the 2020 L2RPN competition, showcased at NeurIPS 2020, had increased technical complexity. It came with a very large discrete and combinatorial action space dimensions offered by the topological flexibilities of the network. It also introduced a realistically-sized network environment along two tracks, representing two realistic real-life network operation challenges; robustness and adaptability (Section

---

1. GridAlive ecosystem https://github.com/rte-france/gridAlive

2. Grid2op package https://github.com/rte-france/Grid2Op

2.3.1). Participation and activity throughout 2020 were steady and there were entries from all over the world. The materials and resources made available to the participants, in particular the Grid2Op framework, are described in Section 2.4.

Throughout the competition, ML approaches showcased continuous robust and adaptable behaviours over long time horizons. This behaviour was not previously observed by the expert systems, or by optimization methods that are limited by computation time. The results are shown in Section 3 with the best methods described in 3.2. After sharing our organizational insights in Section 4, we conclude on further research avenues to build upon.
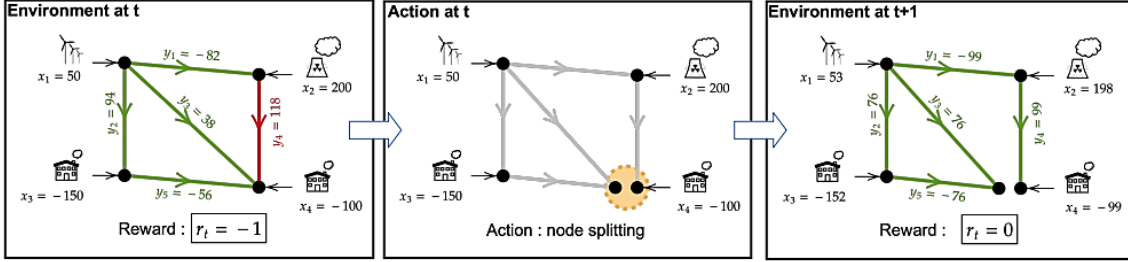

Figure 1: **Power system operation as a complex temporal graph problem:** instead of inferring a given graph properties as in well-known graph problems, we try to find the right graphs over time, among all possible topologies, given the properties that would optimize the flows for safe network operations such as no line overflow.

2. Competition overview

This section details the competition design including the problem, tasks and evaluation and the materials that were made available to the participants. Full details of the competition development can be found in (Marot et al., 2020b).

2.1. Network Operations and Problem Overview

The goal of this competition was to control a power network to maintain a supply of electricity to consumers on the network over a given time horizon by avoiding a blackout. In reality, blackouts are caused by cascading failures of overloaded lines which get automatically disconnected. This can cause loss of power to consumers. Agents should be designed to optimize the **cost of operations** under the safety constraints. The sequential decision making problem is addressed by framing it as a Markov Decision process (MDP) through the lens of RL (Sutton and Barto, 2018).

Objectives and game over condition. Environments are episodic, running at **5-minute** resolution over **a week** (2016 timesteps). The core part of the environment is a power network, with substations (nodes), some containing consumers (load) or production (generation), and interconnecting power lines (edges). The power lines have different physical

characteristics) represented as a **graph**. The industry standard synthetic **IEEE 118** network was used for this competition (Pena et al., 2017). Realistic production and consumption scenarios were generated for the network. The agents can take actions on the network considering the following constraints:

- Events such as maintenance (deterministic) and line disconnections (stochastic and adversarial) can disconnect power lines for several hours.

- Power lines have intrinsic physical capacities, limiting flows beyond. If a power line is overloaded too long, it automatically disconnects. This can lead to cascading failures.

- An agent must wait few hours before reconnecting an incurred line disconnection.

- Additionally, to avoid expensive network asset degradation or failure, an agent cannot act too frequently on the same asset or perform too many actions at the same time.

The Game Over condition is triggered if total demand is not met anymore, that is if some consumption is lost at a substation. More detailed explanation of the physical laws and the physical characteristics of the network can be found in (Kelly et al., 2020) white paper.

Action space. There are two allowable actions: cheap topology changes (discrete and combinatorial actions) that allow the agent to modify the graph of the power network and costly production changes from pre-defined generation (continuous actions). Power production can be modified within the physical constraints of each plant. The final action space has more than $70,000$ discrete actions and a 40-dimensional continuous action space.

Observation space. Agents can observe the complete state of the power network at every step. This includes flows on each power line, electricity consumed and produced at each node, power line disconnection duration etc. After verification of the previously described constraints, each action is fed into an underlying power flow simulator to run AC powerflow (see 2.2) to calculate the next network state. Agent also have the opportunity to **simulate** one's action effect on the current state, to validate their action for instance as operator would do. But the future remains unknown: anticipating contingencies is not possible, upcoming productions and consumption are stochastic.

Reward. There is a standard score function for the organizers to rank each agent (see 2.3.2). But participants were free to design their own reward function for learning.

2.2. Physical power flow simulators

For decades, the power system community has dedicated resources to the development of physical simulators and models with defined benchmarks (Zimmerman et al., 1997), and there are many on the market. These simulators calculate power flow for a given network model and a set of initial physical conditions (production and consumption at each node, and topology). Traditionally in the community, "optimal power flow" (OPF) problems (Babaeinejadsarookolaee et al., 2019) are cast into optimization frameworks which rely on different assumptions, leading to simplifications and tailored simulator implementations.

By casting the problem as a MDP, no strong assumptions about the network dynamics are necessary anymore. The existing complex simulators which solve the physical equations

can be directly leveraged within the Grid2Op framework, letting agents learn through interactions. Aforementioned shortcomings are avoided in L2RPN, the first benchmark that evaluates the continuous time problem while incorporating the complexity. This framework would further apply to upcoming **cyber-physical** power system environments with multiple agents and automaton interacting on top of the historical physical network layer.

2.3. Competition design

2.3.1. Adaptability and Robustness tracks

The 2020 competition was divided into two tracks as depicted in Figure 2, which aimed to solve different aspects of the network operation problem. Participants were given hundreds of years of *chronics* which are time-series production, consumption and maintenance schedules to train on. More details can be found in (Marot et al., 2020b).

Robustness Track: In this track, agents had to operate the network, maintaining supply to consumers and avoiding overloads while an "opponent" takes adversarial actions. This is achieved through targeted line disconnections, as shown on Figure 2 (left), which degrade network operating conditions. The opponent was heuristically designed, disconnecting one of the most loaded lines at random times (Omnes et al., 2020).

Adaptability Track: In this track the ability of agents to cope with different, unseen data distribution within a range of possible distributions was assessed. For training, the participants were given several datasets that were generated from five different energy mix distributions, varying from 10% (1x) to 30% (3x) renewables as shown in Figure 2. These energy mix distributions were known by participants during training. At test time however, the agents had to respond to unknown energy mix distributions within the 1x-3x range.
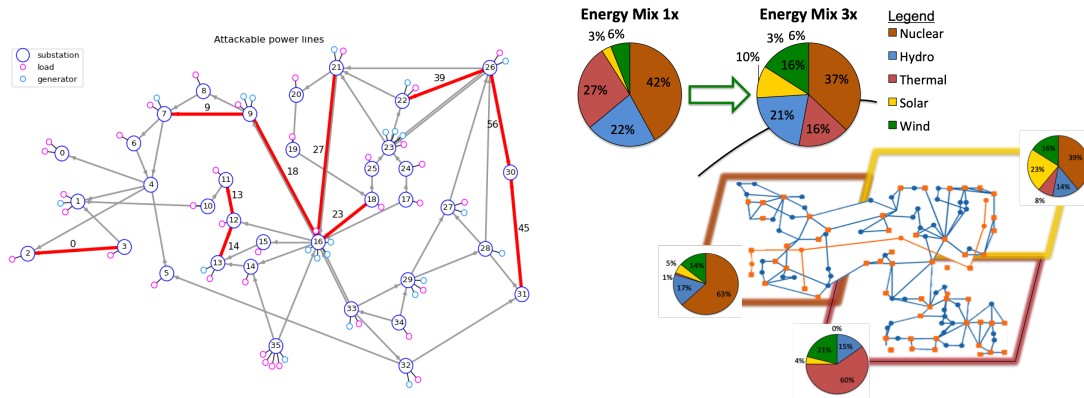

Figure 2: Left - Robustness track top right area of 118 bus network, red lines indicate lines under attack. Right - Adaptability track, pie charts are production mixes with varying renewable integration (from 10% to 30%) and with differences per area.

2.3.2. Evaluation and ranking

Three Phases The competition was divided into three phases. During the warm-up phase, participants could ask any question, some modifications were added to the Grid2op framework based on participants feedback, and new packages could be natively added on the challenge platform. They could already get familiar with the remote submission process, their score appearing over a leaderboard. The evaluation environment was then fixed.

The next phase was the feedback phase, which allowed participants to train their model and assess their performance on an unknown validation dataset, getting some granular feedback on their agent survival time per scenario. Participants could develop and improve their model iteratively and regularly. To avoid over-fitting, the final score was computed during an automatic test phase on a new but similar hidden dataset.

Evaluation Participants were evaluated over 24 independent weekly **test scenarios**, 2 over each month of the year. These undisclosed evaluation datasets were specifically chosen by the competition organizers to be representative of the diverse problems faced by network operators, such as overflow due to high load, or high renewable generation.

The validation and test datasets were drawn from the same statistical distribution, applying the same methodology to both. The **score** per scenario was based on the cumulative network operational cost, further normalized as: [-100,0,80,100] = [initial blackout, do-nothing agent score, scenario completed, additional 20%operational cost optimization].

The leaderboard score was the average score over all 24 scenarios. The running time of the agents on the evaluation platform was eventually limited to 30 minutes, beyond which -100 score was applied. More details can be found in (Marot et al., 2020b).

2.4. Innovative Testbed Framework and Other Competition Resources

Grid2Op L2RPN competitions could use any powersystem simulator to emulate the operational decision making process and cast it as a Markov Decision Process, as explained in 2.2. The most important innovation and the key tool given to participants was the Grid2op python package[2]. This framework is similar to other recent environments such as RLGC (Huang et al., 2019), pymgrid (Henri et al., 2020), gym-anm (Henry and Ernst, 2021) or open modelica microgrid (Heid et al., 2020) in the sense that it allows agents to interact, using the standard "openAI gym" interface (Brockman et al., 2016), with a power network.

One major difference for the L2RPN competitions lies in the **advance action modeling**, which utilizes actions on productions and on the topology of the network. By design, Grid2op is also more **modular**. It allows in particular users to plug in the underlying network simulator model of one's choice, as opposed to other frameworks which depend on one specific simulator. It can in particular get interfaced with **industrial simulators**, such as the ones in PowSybl framework [3], enabling faster reuse of developed solutions on more industrial cases. For the L2RPN competitions, the participants could either use the default simulator embedded in the Grid2op platform which is based on pandapower (Thurner et al., 2018) or they could use a port in C++ to gain computation speedup.

---

3. see https://github.com/powsybl

The Grid2op framework comes equipped with a set of notebooks and extensive technical documentation available online[4]. This ensured that participants could learn and interact with Grid2op relatively easily, even for newcomers in RL or in power network operation. In a post-competition survey (see 4.2), the overall usability of grid2op was rated 3.8 / 5.

Competition Resources Participants could download a "training environment" that they could use as they wished. These environments were made of different generated scenarios, often representing thousands of simulated years of a network (the training environments were different for each track, "Adaptability" and "Robustness" but were identical for each phase: "warm-up", "validation" and "test") as described in 2.3.

Participants had to submit code that could be executed on the evaluation server, which is not an easy process. To make it as easy as possible, participants were also given a "*starting kit*" containing descriptions of the environments and the problem to tackle with some examples of valid submissions and a dedicated notebook to help the debugging process.

L2RPN extra materials [5] To make the competition accessible and attractive to everyone, the problem was thoroughly presented in the "*IEEE Big Data Analytics*" webinars with live QAs, and additional materials were provided. A white paper was first release on ArXiV (Kelly et al., 2020) to advertise L2RPN to as wide an audience as possible. A "sandbox" competition, on a smaller network from L2RPN 2019 competition, was available as a hands-on. The sandbox was particularly helpful at the beginning, but has been less used ever since. Finally, we also open sourced a "baselines" repository that contains code of past winning submission (top 1 and 2 of L2RPN 2019, as well as the open sourcing, during the Neurips competition of the winning approach of L2RPN WCCI 2020). More than a dozen of such "toy models" were available to the participants, though only a small half of the participants used this material during the competition, according to the post competition survey.

3. Outcomes for Participants

This section presents and explores the results of the competition, from final leaderboards to a detailed analysis of agent behaviour and a description of the winning approaches.

3.1. Submission Performance Overview

By the end of the competition, about 15 participants, using different models, achieved scores above 25 (a very decent base score achieved by our Adaptive Expert System baseline (Marot et al., 2018)) on the Robustness Track, within the 30-min computation time budget. This made the Robustness Track highly competitive. In contrast on the Adaptability Track only about half as many participants surpassed the baseline score. This difference could be explained by a smaller network, hence a smaller observation space dimension, for the Robustness track: it might have required less resources for training. An additional interesting factor was the success of the open-submission phase which increased active participation and performance beyond top-3 teams. As shown on the final leaderboards from Table 1,

---

4. see https://grid2op.readthedocs.io/en/latest/
5. Papers, webinar and best team videos can be found at https://l2rpn.chalearn.org/

Table 1: Top Scores in Robustness (left) and Adaptability tracks (right)

Agent	Score	From
rl_agent	59.26	China
binbinchen	46.89	China
lujixiang	44.62	China
djmax008	43.16	China/USA
tlEEmCES	41.38	Germany
jianda	39.68	Singapore
UN_aiGridOp	34.72	Columbia
yzm	33.50	Singapore
Konstantin	30.38	Russia
panda0246	27.36	China
taka	26.73	Belgium

Agent	Score	From
rl_agent	51.06	China
kunjietang	49.32	China
lujixiang	49.26	China
djmax008	42.70	China/USA
TonyS	28.02	USA
taka	27.94	Belgium
yzm	19.76	Singapore
UN_aiGridOp	16.76	Columbia
var764	15.24	USA
PowerRangers	12.86	USA
mwasserer	7.10	Austria

Track	Use RL	Mixed Actions	Graph feat.	Simulate
Robustness	4/5 (8/10)	4/5 (9/10)	2/5 (5/10)	5/5 (10/10)
Adaptability	4/5 (9/10)	3/5 (5/10)	1/5 (1/10)	5/5 (10/10)

Table 2: Overview of preferred approaches for **top-5** (resp. **top-10**) teams. Mixed actions refers to using both continuous and discrete actions.

the 4 top Teams, from Baidu, Huawei, State Grid of China, Geirina, are the same in order on both Robustness and Adaptability Tracks: China definitely made a splash in our competition. While the teams from State Grid of China and Geirina have significant power systems expertise, the performance of Baidu and Huawei was very impressive given their lack of experience in the power sector.

On the Robustness track, team 'RL-agent' from Baidu won by a very comfortable margin. While they reached the 45-point glass ceiling by mid-competition like the two other winning teams, they eventually broke it just 10 days before the end of the competition: they introduced a new successful feature in their approach through evolutionary policy training. This is a sign of methodological breakthrough beyond traditional approaches (see 3.2.1). Furthermore, comparing agents to WCCI competition agents, we highlighted that introducing an opponent during training lead to a more robust and eventually better performing agent (Omnes et al., 2020) with table shown in Appendix 7.1.

On the adaptability track, scores were a lot tighter. The top scores were achieved earlier in the competition (around October 8 from Figure 4). While the final leader board remained uncertain until the end, no new feature was successfully introduced as a last development to make a strong difference. Even if the scores are already good, there might remain more room for improvement on this adaptability problem, especially under higher variance environments with higher renewable integration, as we will see in Section 3.3.

3.2. Summary of Top 10 Solutions and description for Top 3 methods

Table 2 summarizes interesting features used by top ranked participants at test time. On both tracks, 4 out of the 5 top teams used RL agent models. Surprisingly, the second best agent of the Adaptability track, kunjietang, used a simple expert system running simulations over an ordered and limited set of 200 appropriate topological actions. On the Robustness track, tlEEmCES team (ranked 5th) did similarly. This sets bounds on the added value of Machine Learning on those problems. Note that all RL-based agents also run simulations online to select an action among the best proposed ones: they did not totally trust RL only decisions. All teams eventually reduced the action space to a maximum dimension of 1000 unitary actions for exploitation, while dynamically prioritizing them through the learnt model. The best teams often used a combination of discrete topological and continuous redispatching actions: redispatching, despite being costly, was most likely necessary on tough scenarios, especially on robustness track (4/5 teams). All teams further used high level expert rules such as *"do-nothing if no lines loaded close to their limits"*, *"reconnect lines in maintenance or attacked when allowed"*, or more generically *"recover initial configuration when possible"*.

As for the observation, they all used line flow usage rate (variables to control) and connectivity status (topological variables to act on). Consumption and production observations were mostly used by agent playing redispatching actions. Calendar variables such as month, day of the week, hour of the day were surprisingly rarely used, except for winning RL agent and TonyS. Finally, fewer teams used graph related features of this problem, mostly on track robustness: djMax008, jianda and UN aiGridOp used a graph neural network, tlEEmCES used the grid adjacency matrix through an OPF for redispacthing while yzm computed a graph centrality measure feature. Interestingly, Kaist, the winning team from the WCCI competition, did not perform well with their graph neural network approach in this competition, though they did not participate actively. Considering the reward function, RL agent used the simplest possible one: survival duration for the agent. Others used a bit more customized ones, but without much innovations from the ones already available in Grid2op. Reward such as for lujixiang team could be the combination of multiple criteria. But no participants used multi-objective rewards, while available in Grid2Op similarly to safety gym (Ray et al., 2019).

Overall, while action validation through simulation was a must-have, we believe that discovering robust set of actions during exploration, as well as learning an ability to plan for action sequences as for RL agent and binbinchen, made the strongest difference.

3.2.1. Top-3 Team approaches

rl agent - 1st place This team of experienced RL researchers from Baidu won both tracks. They have won similar competitions at Neurips in 2018 and 2019 (Learning To Run and Learning to Move (Zhou et al., 2019)). Relying on their "PARL"[6] framework and the computational resources of Baidu, they were able to train a top performing agent on 5000 cores in 30 minutes, displaying seemingly super-human moves[7] (see Appendix 7.3).

---

6. Available at https://github.com/PaddlePaddle/PARL

7. Video "Advanced Behaviour Analysis of best AI agent" https://www.youtube.com/watch?v=xlqS-CzvMwk

After an action space reduction to 1000 elements with simple expert systems, they initialize a policy parametrized by a feed forward neural network with a few million parameters. Then, they train a policy using evolutionary black box optimization (Lee et al., 2020). As opposed to most standard deep RL strategies, e.g. Proximal Policy Optimization (PPO) (Schulman et al., 2017) or Dueling Deep Q Network (DDQN) (Wang et al., 2016), they didn't rely exclusively on gradients but on genetic algorithm at last to find the "optimal" weights for their policies, enabling them to eventually overcome large variance issues in reward and value estimation. It improved by 10-points their performance when eventually used in their last-week submitted agents.

binbinchen - 2nd place The team from Huawei used an RL agent based on an Actor-Critic model with PPO to play topological actions. An expert-guided exploration phase starts by reducing the action space to 200 actions. The resulting expert agent achieved a score around 40-45, tackling the issue of large action space to learn from. Then they learned a new agent by imitation with a simple feed-forward neural network, performing a bit worse than the expert agent, reaching a score of 40. To better deal with game constraints requiring planning abilities and achieve long-term rewards, they eventually learn a RL agent, whose actor network is initialized from the imitation agent. This agent achieves a score closer to 50, a 10 point increase from imitation learning reflex agent only.

lujixiang - 3rd place This team used a DDQN-based approach (Lan et al., 2020) that was very similar to the winning team from L2RPN 2019, djmax008.

3.3. Results details and discussion

In this section we explore the results achieved to gain insight into the challenges posed by different scenarios and to better understand the behaviour of the top 3 agents.

3.3.1. Impact of Different Scenarios on Performance

Exploring the variation in the performance of agents across the different test scenarios helps characterise the agents and also the factors that make a scenario more challenging. On some scenarios all agents did equally well, there are other scenarios where only one or two agents performed successfully, and finally some scenarios where all agents failed. In the Robustness track, almost all of the 9 unfinished scenarios occurred during winter months, during which the consumption is actually higher leading to a more highly loaded network. In the Adaptability track, the seven unfinished scenarios occur for higher renewable energy mix mix03 (27% renewables) and mix04 (32%renewable). Note that mix04 was slightly out of training distribution range (10-30% renewables), testing for an additional extrapolation capability, which eventually appeared difficult for agents. All scenarios for mix01 (17% renewables) and mix02 (22% renewables) are properly managed by at least one agent. Best agents eventually highlight aposteriori the feasibility frontiers, pushing forward our knowledge about the power network flexibilities. Agent episode traces are available for inspection through our Grid2Viz study tool[8]. More details and survival time heatmaps can also be found in Appendix 7.2.

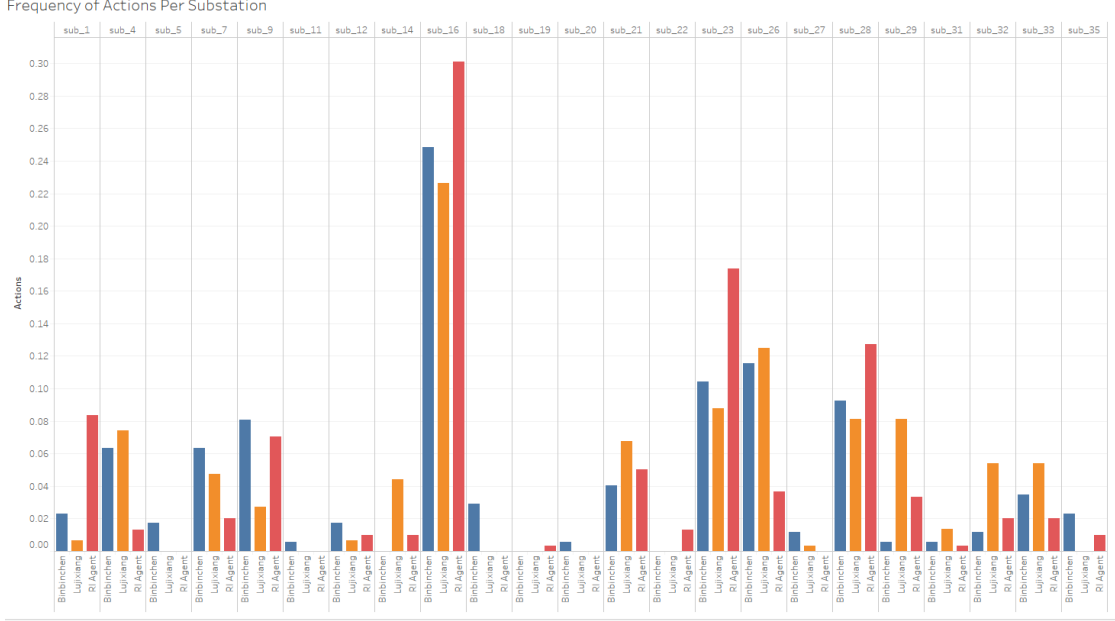8. Grid2viz tool for best agent study https://github.com/mjothy/grid2viz

Figure 3: Frequency of unitary substation re-configurations utilized by the top 3 agents across all scenarios.

3.3.2. Agent Behaviour analysis

The problem posed by this challenge is extremely complex and agents deploy different strategies to respond to the same scenarios. In this section we attempt to gain insight to these strategies by analysing the behaviour of the top 3 agents on Robustness track across three key dimensions, **action diversity** in the number of substations controlled and number of actions taken and also **sequence length of the re-configurations** which provides insight into the agent's ability to combine actions in sequences.

Action diversity   Assessing the agents approach to the substation topology re-configurations reveals distinct patterns of behaviour. First we note that the agents RLagent and lujixiang utilise significantly more substation re-configurations across all scenarios (namely 299 and 296, respectively). Agent binbinchen performed about half as many (namely 173). Next, Figure 3 shows how the actions are distributed across the different substations. While performing the least amount of actions binbinchen (blue) still engages more substations (namely 20) than RLagent (red, 18) and lujixiang (orange, 16), all demonstrating an interesting action diversity. The substations 16, 23, and 28 represent hubs in the power network and, hence, are often used by all agents. Other crucial substations are only relevant to two of the three agents (e.g. substations 4, 26, 29). Finally, both binbinchen and RLagent engage substations that are not used by the other agents (substations 5, 11, 18, 20 and 19, 22, respectively).

Sequence length of reconfigurations   All agents predominately (for more than 70% of the action sequences, not shown) perform a single substation re-configuration and then
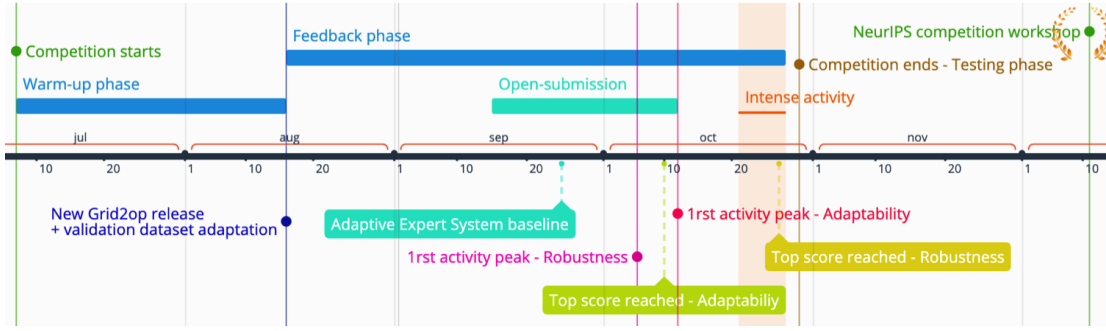
11

Figure 4: Competition timeline displaying different phases and interesting events

do nothing again. Agent binbinchen is especially concise with more than 80% of its action sequences being of length 1. Agent RLagent on the other hand is most pronounced when it comes to executing sequences of length 2-3 (more than 20%). Finally, agent lujixiang has a tendency to create long sequences of up to 15 consecutive substation re-configurations (involving up to 7 different substations). The longest sequences of the agents binbinchen and RLagent have a length of 5 and 6, respectively.

In summary, the 3 agents display distinct patterns of Behaviour which demonstrates how the number of controlled substations and the length of action sequences can be balanced in different ways. Agent binbinchen tends to perform a small amount of actions executed in short sequences but at the same time involves many substations. In contrast, lujixiang operates on a smaller set of substations but performs more actions organised in relatively long sequences. The Behaviour of RLagent is intermediate by engaging a moderate amount of substations in mid-sized sequences, which might be a preferred Behaviour to operators.

4. Organizational Outcomes

The competition eventually ran smoothly on Codalab competition platform through our different phases displayed on Figure 4 timeline. We review what contributed to a successful and dynamic competition, while also giving recommendations for possible improvements.

4.1. Open-submissions, 2 tracks and 3 phases for a dynamic competition

Open-submissions: We aimed at increasing overall participation and interest, by giving everyone a chance to get a prize for their efforts while making their submission public for others. While almost no open-submission happened in the first two weeks, it became successful after we released a new Adaptive Expert System baseline (Marot et al., 2018)), paving the way and increasing participant awareness. Setting an affordable minimum score threshold of 10, 5 best in time and 5 best in score teams won 200$ and 400$.

There was an increase of activity and in the number of new participants by the end of open-submission period. First high activity peaks (over 40 submissions a day) occurred in this period, while we usually see such peaks mainly during the last week of competition. The winning open-submission teams were logically different from the leading teams who did not want to reveal their best agent. The best open-submissions actually combined

successfully different parts from other open-submissions, making the competition more and more competitive, bridging some gap with the four leading teams. This also highlights the potential for collaboration in competitions, which could be further explored.

2 tracks: The group's previous experience organising competitions made it possible to streamline the set up process. This allowed two separate tracks to be designed instead of one, on equally important but orthogonal problems for real-world power network, without significant additional investment of resources. It hence let participants chose one were they would be more comfortable or capable with and best compete. While decoupling the two problems makes it more workable and clearer for participants, it also let some team leverage their investment on one (about 2 months), to compete in a shorter time frame with a performant model on the other (only 2 to 3 weeks). Except for the winning teams, teams were mostly different on both tracks. Having two leaderboards, we saw overall an increase of active participation and interest compared to a single track. Both tracks ended at the same time in our competition. As peaks of activities occur near the end of the competitions, it might be interesting to actually end a second track two or three weeks after the first one.

Finally, our new warm-up phase was appropriate to get participants engaged, keep improving the framework for competition needs, and adjust validation set difficulty.

4.2. Effectiveness of Competition Materials

We evaluated how useful the competition materials were for the participants, based on what we have learnt through our Discord forum and competition survey. Among the 64 unique teams that appeared on the leaderboard and eventually competed, 16 filled out the survey form made of 35 questions through 4 categories "*competition in general*", "*Grid2op and the competition ecosystem*", "*Codalab, making submissions and competition progression*" and "*Organization and communication during the competition*".

Overall, all participants seemed to have enjoyed the competition: they rated it 8 / 10 and the Discord forum was greatly appreciated (4.7 / 5) which we recommend to other organizers. We indeed think it helps the participants get engaged and get customized help much more quickly: more than 4.700 messages were posted by both participants and organizers. It also allowed us to better understand the needs of the participants and adapt the evaluation servers to their needs with desired software.

As anticipated the submission of code produced some challenges. The most difficult part, for participants, was to provide easily reproducible code that would run on evaluation servers. The question "*How easy was the submission process ?*" was the one with the lowest but yet encouraging average value: 3.6 / 5. Concerning the competition's difficulty, surprisingly only 1 participant found it difficult among the respondents (only answer below 3)[9]. Despite this potential bias, we believe this to be representative of the improvement in the submission process, the "getting started" notebooks, and the help provided in the Discord forum. The improvement in the submission process is reflected by the number of submissions made during this competition compared to the previous ones. Indeed, more than 2700 independent submissions were made during the feedback phase, as opposed to

---

9. Though this statement is probably slightly biased, as probably the people who experienced difficulties did not engage until the end of the competition or were less likely to take the post-competition survey

approximately $200 \sim 300$ for the 2019 L2RPN competition and $700 \sim 800$ for the 2020 WCCI L2RPN competition.

4.3. Soundness of Evaluation Methodology

Our scoring over 24 scenarios proved to be robust for the two tracks as the best teams were ranked in same order with similar scores between feedback and test phases, except for one team. The exception was due to overfitting by UN-aiGridOperator Team. Winning the open-submission on the feedback phase, they suffered from overfitting, with a 10-point loss on the test phase. This is one more example of the soundness of having two distinct validation and test sets when evaluating models. The normalized scores per scenarios were more interpretable compared to our first competitions. It helped us better appreciate the performance and Behaviour of agents on each scenarios, giving richer feedback to the participants as well. It also let us compare results on both tracks.

However there were some opportunities for improvement and lessons learned. A new criteria to filter out hot start scenarios would be useful as analyzed in Appendix 7.2. In addition, our normalization function around 0 defined by the do-nothing survival time was sometimes a bit too harsh, with agent losing up to $20/100$ points in score for only few timesteps. This could be softened to be more representative of the agent performance when it does not do very badly. Finally, $20/100$ points per scenario would have been given to agents for continuously optimizing the cost of energy losses, beside surviving. But none won those points, making us wonder how complex this additional task is and what should be the right level of incentives for participants to start tackling it beyond safety. This multi-objective evaluation could be framed differently in the future.

5. Conclusion

The L2RPN NeurIPS competition has demonstrated how encouraging collaboration between the AI and power systems communities can lead to the advancement of novel and innovative AI methods for one of the most significant and challenging problems in the power sector, grid operation. The winning team RlAgent was a team of researchers from Baidu, a tech company, with no domain expertise in power systems. This highlights the value in making tools like the Grid2Op platform available, which lower the barriers for entry in this domain, and the benefit of running global competitions, which attract the interest of leading AI researchers. As well as being a problem with valuable real-world applications the L2RPN challenge is also pushing the state of the art of AI forward as a uniquely challenging problem which involves graph topology, complex physics and high dimensional action spaces which have non-local effects throughout the network. Model-based approaches such as physics-informed ones could be relevant in that regard to make further progress.

Yet, the behaviour demonstrated by the best agents was very encouraging with interesting successful long action sequences even under strong perturbations, showcasing some desired robustness. It showed that it could augment knowledge of grid flexibility beyond human capability, which would already be a valuable application for operators when applied on their grid. It needs to be highlighted that in this problem setting identifying the 'optimal' response that an agent should take is almost impossible due to the combinatorial nature of the problem. Deeper analysis of agent behaviour will have revealing insights

for power systems operators. Solutions could be extended to capture further complexity; currently no continuous power loss optimization was performed by agents other than satisfying safety considerations: this would be a highly impactful next addition. As the power system evolves to meet the decarbonisation targets it will become increasingly complex for operators who will require new tools and support in their decision making. An additional research avenue is trustworthy human-machine interactions as we envision AI assistants for operators (Marot et al., 2020c) rather than a fully autonomous system: it comes as the next research direction showcased in our new "L2RPN with Trust" competition[10].

It is hoped that this challenge opens up further fruitful and important collaborations between the power system community, which has a host of complex decision making problems which need solutions, and the AI community who can benefit from the opportunity to apply Reinforcement and Deep Learning methods to real-world problems. Through the growing open-source GridAlive ecosystem, researchers are encouraged to share new agent baselines to be benchmarked and studied, as well as to build new relevant environments. It is hoped that new competitions will develop given their attractive catalyst effect for research, and that the insights discussed in this paper will prove useful to new organizers. Lastly, L2RPN interactive materials are a valuable resource for education and teaching.

6. Acknowledgements

References

Sogol Babaeinejadsarookolaee, Adam Birchfield, Richard D Christie, Carleton Coffrin, Christopher DeMarco, Ruisheng Diao, Michael Ferris, Stephane Fliscounakis, Scott Greene, Renke Huang, et al. The power grid library for benchmarking ac optimal power flow algorithms. *arXiv preprint arXiv:1908.02788*, 2019.

Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning*, pages 507–517. PMLR, 2020.

Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

Noam Brown and Tuomas Sandholm. Safe and nested subgame solving for imperfect-information games. In *Advances in neural information processing systems*, pages 689–699, 2017.

---

10. "L2RPN with Trust" https://icaps21.icaps-conference.org/Competitions/

Jeremy Cohen. How google's self-driving cars work, 2020. URL https://heartbeat.fritz.ai/how-googles-self-driving-cars-work-c77e4126f6e7m.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

Gabriel Dulac-Arnold, Daniel J Mankowitz, Nir Levine, Cosmin Paduraru, Jerry Li, Sven Gowal, and Todd Hester. Challenges of real-world reinforcement learning: Definitions, benchmarks & analysis. 2021.

Mevludin Glavic, Raphaël Fonteneau, and Damien Ernst. Reinforcement learning for electric power system decision and control: Past considerations and perspectives. *IFAC-PapersOnLine*, 50(1):6918–6927, 2017.

Stefan Heid, Daniel Weber, Henrik Bode, Eyke Hüllermeier, and Oliver Wallscheid. Omg: A scalable and flexible simulation and testing environment toolbox for intelligent microgrid control. *Journal of Open Source Software*, 5(54):2435, 2020. doi: 10.21105/joss.02435. URL https://doi.org/10.21105/joss.02435.

Gonzague Henri, Tanguy Levent, Avishai Halev, Reda Alami, and Philippe Cordier. pymgrid: An open-source python microgrid simulator for applied artificial intelligence research. *arXiv preprint arXiv:2011.08004*, 2020.

Robin Henry and Damien Ernst. Gym-anm: Reinforcement learning environments for active network management tasks in electricity distribution systems. *arXiv preprint arXiv:2103.07932*, 2021.

Qiuhua Huang, Renke Huang, Weituo Hao, Jie Tan, Rui Fan, and Zhenyu Huang. Adaptive power system emergency control using deep reinforcement learning. *IEEE Transactions on Smart Grid*, 11(2):1171–1182, 2019.

Chris Gamble Jim Gao. Safety-first ai for autonomous data centre cooling and industrial control, 2018. URL https://deepmind.com/blog/article/safety-first-ai-autonomous-data-centre-cooling-and-industrial-control.

Adrian Kelly, Aidan O'Sullivan, Patrick de Mars, and Antoine Marot. Reinforcement learning for electricity network operation. *arXiv preprint arXiv:2003.07339*, 2020.

Łukasz Kidziński, Carmichael Ong, Sharada Prasanna Mohanty, Jennifer Hicks, Sean Carroll, Bo Zhou, Hongsheng Zeng, Fan Wang, Rongzhong Lian, Hao Tian, et al. Artificial intelligence for prosthetics: Challenge solutions. In *The NeurIPS'18 Competition*, pages 69–128. Springer, 2020.

Tu Lan, Jiajun Duan, Bei Zhang, Di Shi, Zhiwei Wang, Ruisheng Diao, and Xiaohu Zhang. Ai-based autonomous line flow control via topology adjustment for maximizing time-series atcs. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2020.

Kyunghyun Lee, Byeong-Uk Lee, Ukcheol Shin, and In So Kweon. An efficient asynchronous method for integrating evolutionary and gradient-based policy search. 2020.

Zhongyu Li, Xuxin Cheng, Xue Bin Peng, Pieter Abbeel, Sergey Levine, Glen Berseth, and Koushil Sreenath. Reinforcement learning for robust parameterized locomotion control of bipedal robots. *arXiv preprint arXiv:2103.14295*, 2021.

A Marot, B Donnot, S Tazi, and P Panciatici. Expert system for topological remedial action discovery in smart grids. *IET Digital Library*, 2018.

Antoine Marot, Benjamin Donnot, Camilo Romero, Balthazar Donon, Marvin Lerousseau, Luca Veyrin-Forrer, and Isabelle Guyon. Learning to run a power network challenge for training topology controllers. *Electric Power Systems Research*, 189:106635, 2020a.

Antoine Marot, Isabelle Guyon, Benjamin Donnot, Gabriel Dulac-Arnold, Patrick Panciatici, Mariette Awad, Aidan O'Sullivan, Adrian Kelly, and Zigfried Hampel-Arias. L2rpn: Learning to run a power network in a sustainable world neurips2020 challenge design. 2020b.

Antoine Marot, Alexandre Rozier, Matthieu Dussartre, Laure Crochepierre, and Benjamin Donnot. Towards an ai assistant for human grid operators. *arXiv preprint arXiv:2012.02026*, 2020c.

Sharada Mohanty, Erik Nygren, Florian Laurent, Manuel Schneider, Christian Scheller, Nilabha Bhattacharya, Jeremy Watson, Adrian Egli, Christian Eichenberger, Christian Baumberger, et al. Flatland-rl: Multi-agent reinforcement learning on trains. *arXiv preprint arXiv:2012.05893*, 2020.

Loïc Omnes, Antoine Marot, and Benjamin Donnot. Adversarial training for continuous robustness control problem in power systems. *arXiv preprint arXiv:2012.11390*, 2020.

Ivonne Pena, Carlo Brancucci Martinez-Anido, and Bri-Mathias Hodge. An extended ieee 118-bus test system with high renewable penetration. *IEEE Transactions on Power Systems*, 33(1):281–289, 2017.

Alexander M Prostejovsky, Christoph Brosinsky, Kai Heussen, Dirk Westermann, Jochen Kreusel, and Mattia Marinelli. The future role of human operators in highly automated electric power systems. *Electric Power Systems Research*, 175:105883, 2019.

Alex Ray, Joshua Achiam, and Dario Amodei. Benchmarking safe exploration in deep reinforcement learning. *arXiv preprint arXiv:1910.01708*, 2019.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. Mastering chess and shogi by self-play with a general reinforcement learning algorithm. *arXiv preprint arXiv:1712.01815*, 2017.

Medha Subramanian, Jan Viebahn, Simon Tindemans, Benjamin Donnot, and Antoine Marot. Exploring grid topology reconfiguration using a simple deep reinforcement learning approach. *arXiv preprint arXiv:2011.13465*, 2020.

R. S. Sutton and A. G. Barto. Reinforcement learning an introduction. *MIT press*, 2018.

Leon Thurner, Alexander Scheidler, Florian Schäfer, Jan-Hendrik Menke, Julian Dollichon, Friederike Meier, Steffen Meinecke, and Martin Braun. pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems. *IEEE Transactions on Power Systems*, 33(6):6510–6521, 2018.

Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. Grandmaster level in starcraft ii using multi-agent reinforcement learning. *Nature*, 575 (7782):350–354, 2019.

Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas. Dueling network architectures for deep reinforcement learning. In *International conference on machine learning*, pages 1995–2003. PMLR, 2016.

Deunsol Yoon, Sunghoon Hong, Byung-Jun Lee, and Kee-Eung Kim. Winning the l2rpn challenge: Power grid management via semi-markov afterstate actor-critic. In *International Conference on Learning Representations (ICLR)*, 5 2021. URL https://openreview.net/forum?id=LmUJqB1Cz8.

Bo Zhou, Hongsheng Zeng, Fan Wang, Yunxiang Li, and Hao Tian. Efficient and robust reinforcement learning with uncertainty-based value expansion. *arXiv preprint arXiv:1912.05328*, 2019.

Ray D Zimmerman, Carlos E Murillo-Sánchez, and Deqiang Gan. Matpower: A matlab power system simulation package. *Manual, Power Systems Engineering Research Center, Ithaca NY*, 1:10–7, 1997.

7. Appendix

7.1. Comparison of WCCI and NeurIPS competition agents

WCCI and NeurIPS competitions differ only by the introduction of an opponent. This table 3 is taken from our paper (Omnes et al., 2020) and highlights differences in performance when the opponent was available during training or not. Not only are agents stronger on the NeurIPS Robustness competition when they were given the opponent for training, they also outperform the WCCI agent on the WCCI competition. Plus, while the scores for the NeurIPS agents are only a little bit lower on the NeurIPS competition, which is expected since that competition uses an opponent and is thus harder to deal with, the performance of WCCI agent zenghsh3 (actually from Baidu as well) drops much more dramatically when confronted to the opponent.
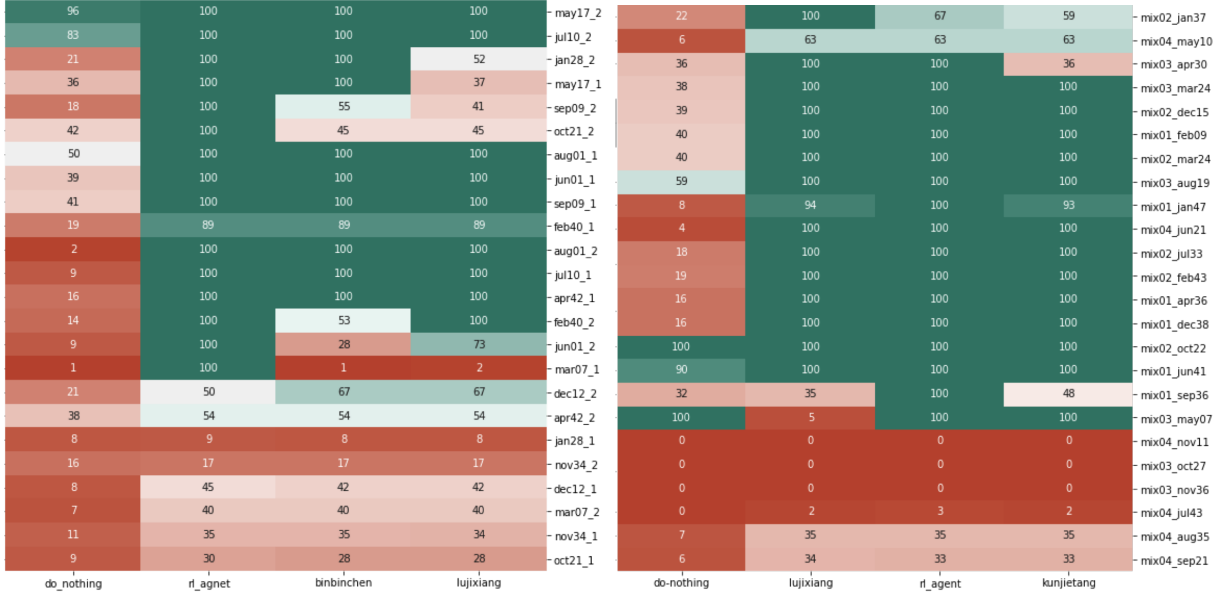
Figure 5: Survival time percentage heatmaps (red for early blackout and green for complete episode) for do-nothing agent baseline and top 3 agent, over the 24 weekly test scenarios for robustness (left) and adaptability (right) tracks

Table 3: Agents results on the WCCI and NeurIPS competitions

| Agent | From competition | Results | |
		WCCI	NeurIPS
rl_agnet	NeurIPS	71.21	61.05
binbinchen	NeurIPS	63.97	52.42
lujixiang	NeurIPS	73.73	45.00
zenghsh3	WCCI	58.21	19.10
reco_powerline	Baseline	25.75	10.76
do_nothing	Baseline	0.00	0.00

7.2. Detailed results over scenarios

We discuss the feasibility of unfinished scenarios looking at the survival time heatmaps of Figure 5. For robustness track, we find that a set of scenarios (Jan 28.1 + nov34.1 + apr 42.2 + oct 21.1) looked tough but feasible at the time of gameover using only topological actions, another set (Feb40.1 + March 07.2 + dec12.2) would require combined redispatching type of action, while a last set (Nov34.2 + dec 12.1) looked eventually nearly impossible without new flexibilities (voltage control, load shedding) not available during the competition. For Adaptability track, three scenarios (mix04_nov11, mix03_oct27 and mix03_nov36) actually

19

appeared infeasible because of a harsh start with many already existing overloads to solve without possible anticipation. Other mix04 scenarios appear quite tough but possible, all presenting a high renewable production evacuation issue at substation 61, sometimes combined to simultaneous overloads in the Southern part of the network. Without any ground-truth, it is often hard apriori to know for some cases if they will indeed be feasible as we have seen here, but this can be refined aposteriori.

7.3. Hard Scenario analysis on Robustness track

Scenarios where the best agent survive longer than other agents, but still fail later, are interesting ones to study from a Behaviour perspective. We here discuss our analysis on such a scenario dec12.1. RL_agent is eventually the only one to survive on December 11th over a strong attack at 6am against a very high voltage line 26-30, as shown on Figure 6. This line was supplying a fair amount of consumption located at the top right of the network. This large powerflow is therefore redipatched over smaller capacity lines connected to this area. Overloads appear on some of these lines, without much margin left on others: the flows needs to be balance timely and precisely, making the task quite complex. Over this 4-hour short period of time, RL_agent execute lots of actions (20), through sometimes impressive complex 3 to 5 combined action sequences. It successfully adapts to new overloads appearing because of a continuously increasing consumption up to 10am. As combined-action depth superior to 3 is hard to study for human operators, some sequences showcased by the agent can be today considered **super-human**.

Ahead of the competition, we anticipated that "it remains largely unknown today how much flexibility an existing network topology can eventually offer when controlled widely at a high frequency because of its non-linear combinatorial complexity". This result shows that AI could indeed help operators augment their knowledge and use of network topology flexibility in the future. Conversely, RL_agent eventually fails over a new attack that a human could most probably have solved. AI can hence be better regarded as an **assistant** for the operator here rather than a fully autonomous agent in charge of the network.
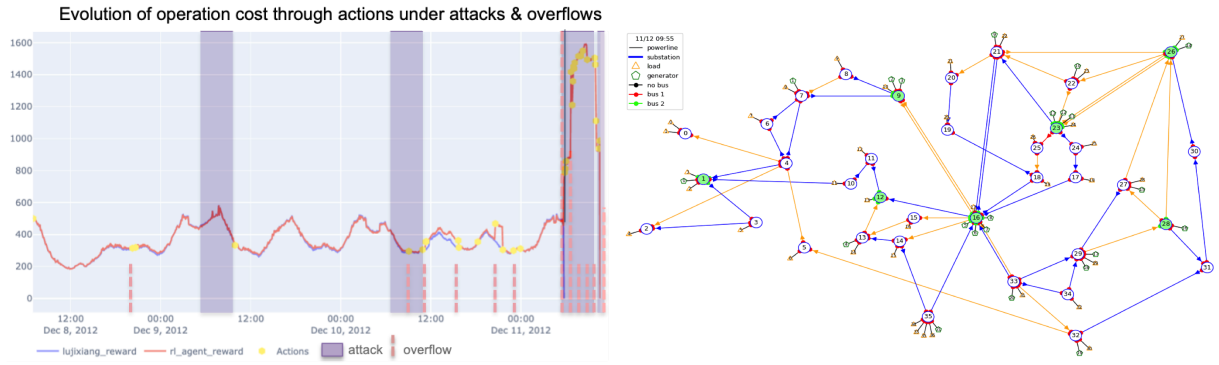
Figure 6: Network operation cost over time (left) for best agents RLagent and lujixiang in scenario dec12.1, overcoming highlighted periods of attack and overflows. Only RLagent manages the third and hard attack by playing very effectively complex sequence of actions, sometimes in a superhuman fashion, over 7 substations (highlighted on the right). It nevertheless fails after a new attack a bit later.