Paris Region AI Challenge for Energy Transition

Low-carbon Grid Operations

April 2023

Contents

1 Preliminaries

1.1 Summary

This document presents a description of the Paris Region AI Challenge for Industry 2023 carried out in collaboration between the Île-de-France region and

RTE (Electricity transport network) with the support of Paris-Saclay University, the ASTech and Systematic clusters, and Startup Inside.

RTE's mission is to transport energy in the form of electricity on long-distance power lines. This mission must be carried out while maintaining the population and equipment safety, which requires monitoring lines at all times to avoid overloading them and not leading to blackouts. RTE is currently deeply engaged in the Energy Transition up to 2050 to massively integrate new renewable energies with the condition to keep a system that can be robustly operated. Changing the energy mix is essential if we want to generate carbon-free electricity. However, it also poses other problems: the intermittent nature of renewables, their unequal geographical distribution, and new uses of electrical power. RTE has to accommodate these new means of power generation and consumption.

The objective of this challenge is to create for the dispatchers of the power grid (see pictures of a dispatching room 2)) a near real-time assistance module offering recommendations for strategies aimed at safely managing overloads on the electrical lines. In this year's competition, a simulation environment allows an artificial agent to act on the power grid scenarios at a 5-minute time step resolution to learn how to operate it and further to be evaluated on its ability to use robust strategies over time, avoiding any black-out. An expectation is also to use this agent as an assistant, integrating trust considerations for the human dispatcher. Finally, this assistant must favor strategies to make the most of the renewable energies installed by limiting the emergency redispatching call for thermal power plants emitting greenhouse gases.

1.2 Confidentiality

This document is not subject to any form of confidentiality and may therefore be distributed freely.

1.3 Disclaimer on document

The information and data contained in this document are published for information only and are not contractual. RTE declines all responsibility for any errors or inaccuracies in the diagrams or explanations. These information may be subject to change without notice.

1.4 Authors and contributors

Antoine Marot, Laure Crochepierre, Karim Chaouache and Benjamin Donnot from RTE as well as Adrien Pavao and Isabelle Guyon from Paris-Saclay University have all contributed to writing this document.

We thank Clément Goubet and Jerôme Dejaegher from RTE as well as Olivier Pietquin from Google and Pr. Madeleine Gibescu at Universiteit Utrech, for their kind review and feedbacks.

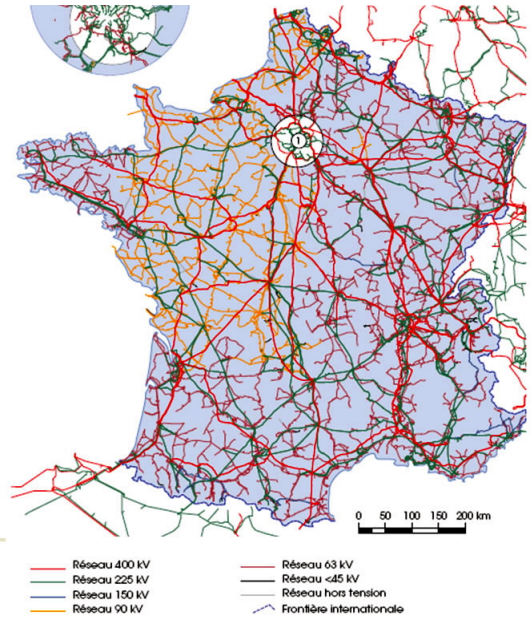

Figure 1: Map of RTE transmission power grid

2 RTE - The French Transmision Grid Operator

Réseau de Transport d'Electricité (RTE) has been the operator of the French power transmission grid since its creation in the year 2000 and continuously fulfills its public service mission for which it is responsible. RTE is the largest European operator in its field with nearly 106,000 km of high and very high voltage lines (see figure map 1), interconnected with the powergrids of its European counterparts within a large European power system.

Its role goes well beyond what the transmission of electricity evokes. Since electricity can only be stored in limited volumes, it must be consumed as soon as it is produced. At the heart of the electrical system, this role gives us first-rate missions:

- Provide everyone, 24 hours a day, 7 days a week, 365 days a year, in France and Europe, with access to an economical, safe, and clean power supply;
- Support and accelerate in the energy transition by welcoming renewable energies and optimizing their contribution while informing public decisions;
- Promote the development of the territories' industrial fabric and participate in French companies' competitiveness.

The challenges of energy transition and the Europe of Energy have led RTE to initiate a profound change. Faced with an environment in the throes of

technological, economic, and social upheaval, RTE must continue its transformation in order to respond to European, national, and territorial challenges by anchoring the performance of its model.

Backed by its powergrid and invested in its public service mission, which is vital for the country and the life of its citizens, RTE works every second to guarantee long-term access to carbon-free electricity. From the European Union to the French territories, the ambitions displayed in terms of the energy transition are considerable. They will lead to profound changes in the electricity sector as a whole: development of renewable energies, increase in exchanges between European countries, new consumer behavior, self-consumption, emergence of new uses, development of electricity storage, etc. These changes are also integrated into the consideration of a technological and digital revolution: new forms of communication, dematerialization, artificial intelligence, geolocation, etc.

In 2021, RTE's turnover amounted to €5,254,036,000. The group has 9,500 employees. More information is available on the website.

3 Energy Transition and Low Carbon Grid Operation Context

In its action to tackle climate change [1], RTE must achieve the objectives of the PPE (Multiannual Energy Program), cutting by half emissions due to the production of electricity by 2035), and of the SNBC (national low carbon strategy) for carbon neutrality in 2050. For this objective to be reached, a massive integration of intermittent renewable energy (solar and wind) is necessary. Energy Futures scenarios [2] indicate, for example, a majority share $> 50\%$ to reach in the mix. And this integration will have to be done at a very sustained pace: there is an urgent need to act whatever the scenario.

This new variability will lead to major challenges in terms of operating the power grid, which will have to be more flexible and responsive, with more complex, rapid, and numerous decisions for dispatchers [12]—always ensuring the essential function of a power grid: supplying consumers while avoiding any blackout, despite this increased complexity. To get an idea of the cost of a blackout, a simulation of a one-hour [3] blackout on March 8 at 2 p.m. in the Paris region could cost around 150 million euros.

An assistant [13, 19] thus becomes key to supporting dispatchers in these decisions within this new environment. New dispatching rooms for the RTE power grid are also currently being deployed, starting with the inauguration of the Paris one in 2023 (see illustrations 3 and 4) (the first in 20 years!) to allow more efficient operation of the grid and meet the challenges facing RTE. This marks a new era for the operation of the grid.

[1] RTE action to tackle climate change https://www.rte-france.com/rte-en-bref/nos-engagements/laction-de-rte-face-au-changement-climatique

[2] Energy Futures 2050https://www.rte-france.com/analyses-tendances-et-prospectives/bilan-previsionnel-2050-futurs-energetiques

[3] Blackout simulator https://www.blackout-simulator.com/

Figure 2: Control Room of the grid today

Figure 3: New control rooms opening in 2023

Figure 4: New operator's desk in new rooms

The integration of these renewable energies, and therefore achieving the objectives to act in the face of climate change, will be a success only if we can continue to operate the grid under high safety conditions, providing operators with these new tools accordingly [6]. In particular, a trusted assistant can recommend strategies to deal with electricity congestion on powerlines. This use case has thus been listed in the AI4Climate report[4] from COP 26 in the UK in 2021.

4 A Dispatcher Assistant: making recommendation with anticipation

Today, human operators are working in real-time from control rooms to optimize the power flows on electrical lines, handle maintenance and new equipment integration, react rapidly to unplanned outages, and, most importantly, avoid (very costly) blackouts. More details about their role and tasks can be found in [17]

Today's operations They are highly trained engineers as their job requires thorough studies, careful planning, and complex decision-making processes rather than simply reproducing pre-established patterns. They heavily rely on simulation tools coupled with real-time and forecast data. But they have little decision-making support tools, such as assistants. When they need to solve a problem, they mostly manually explore solutions and validate their decision in their simulation tool. They can modify the line connectivity on the grid to reroute power flows but also modify some production, limit consumption by a few percent, or even use battery storage today to change the power flows on the grid. This is a large set of possible flexibilities, among which they have to identify the effective ones in a given context. Yet they operate mainly with experience and manual simulation to determine relevant remedial actions.

Tomorrow's sequential operational problem As the grid gets pushed towards its limits, decisions become more numerous and a lot more interdependent. Solutions should not only be effective at one single point in time but over a larger time horizon. Often an action will be beneficial for some issues, but will create more risks elsewhere: there is a permanent trade-off on a grid with a fixed capacity. Actions should also be implemented with the right anticipation given their adequate activation time: switching line connectivity is quick, but starting production can sometimes take a few hours. So decisions will need to indeed consider the full underlying planning problem of power grid operations. This is a continuous sequential decision-making problem.

[4] AI4Climate COP 26 report https://www.gpai.ai/projects/climate-change-and-ai.pdfs

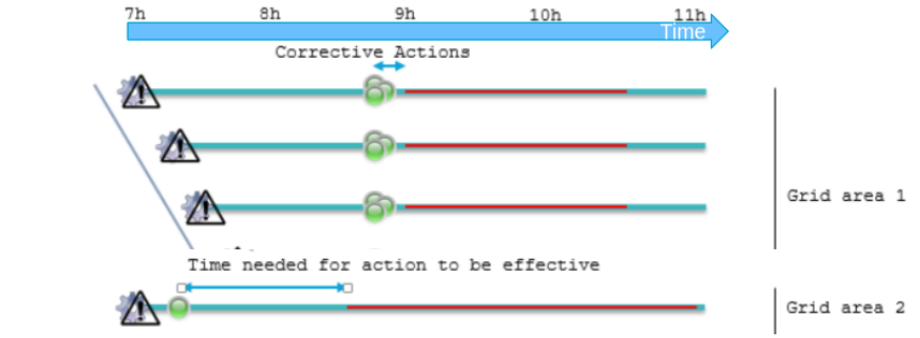

Figure 5: A simple scenario where contingencies are anticipated in multiple parts of a grid with resulting congested periods (in red). Several remedial actions (green dot) with different setup duration (blue arrow) are possible. Choices have to be made with anticipation and coordinated.

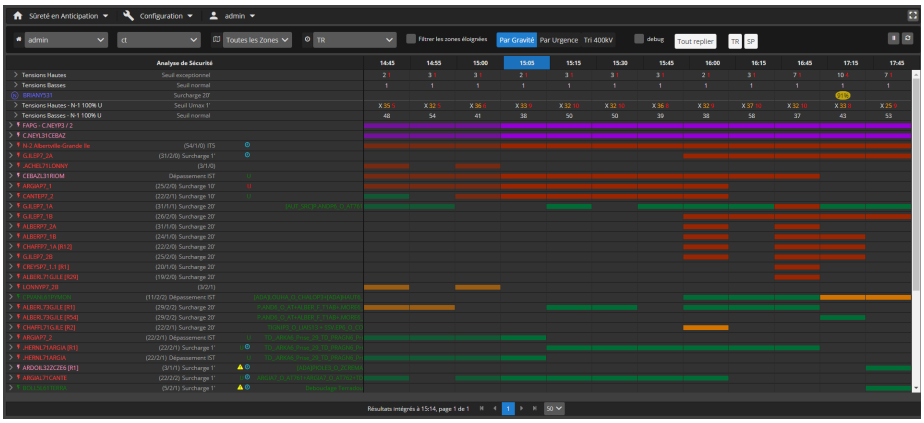

Figure 6: A screenshot of SEA (safety by anticipation) supervision tool. Each row relates to a contingency of interest, and columns are different forecasted time horizons considered. Indicators are green if no issue remains, orange if no issue remains but the effective strategy is not a preferred one, and red if no selected strategy works. An empty indicator means no issue is forecasted.

New tools for recommendation Operators already have some latest (but non-AI) tools that simulate for every upcoming hours over the day the possible consequences of a contingency, such as an unexpected outage on a power line. In case of resulting congestion, this tool allows to simulate automatically 2 to 6 preferred remedial action strategies a priori, with green, orange, and red indicators (see Figure 6). Sometimes operators already have to iterate among a list of 100 possible strategies. As system context changes more rapidly, selected

strategies are not always effective on a new situation, and operators have to iterate among their list to find effective contextual solutions.

In comparison, the assistance recommendation module will bring new capabilities such as :

- taking into account a larger grid context and a time horizon to make more effective recommendations
- providing recommendations on demand regarding new issues to consider in upcoming hours or close to real-time. This offers a lot more interactivity without heavy simulation iterative loops
- indicating its confidence in the effectiveness and robustness of its recommendations, analogous to green and red indicators

In terms of time horizon, the next two-hour horizon can be considered as the operational window, that is the execution phase where it is harder to take time for more studies. Beyond two hour horizon, this is the anticipation phase, during which issues can be studied and related strategies and remedial actions prioritized: this is a configuration phase of an operational plan. The solution that will be developed for real operations should become a game changer within the operational window and later improve the anticipation phase as well.

5 L2RPN problem setting

In this section, we first recall the objective of the formulated challenge. We then formalize the decision-making process and further describe its features.

5.1 Objective

We propose a challenge that will test the abilities of artificial agents at horizon 2030-2035 energy mixes on the way towards 2050 scenarios [1]. The goal is to control electricity transmission in power networks while pursuing multiple objectives: meeting the production/consumption balance, minimizing energy losses, keeping people and equipment safe, and, above all, avoiding catastrophic blackouts. Blackouts happen when a large portion of consumers cannot be supplied anymore because the system is too unstable or because a cascading failure of powerlines happens and no safe electrical path remain to transport electricity to them. Recovering from a blackout often takes at least a few hours, if not days sometimes.

The importance of this application not only serves as a goal in itself but also aims to advance the field of Artificial Intelligence (AI) known as Reinforcement Learning (RL), which offers new possibilities to tackle control problems. In particular, various aspects of the combination of Deep Learning and RL (*Deep Reinforcement Learning*) remain to be harnessed in the domain of electric power networks. This challenge belongs to a series started in 2019 under the name "Learning to Run a Power Network" (L2RPN) (see also Figure 7). In this new

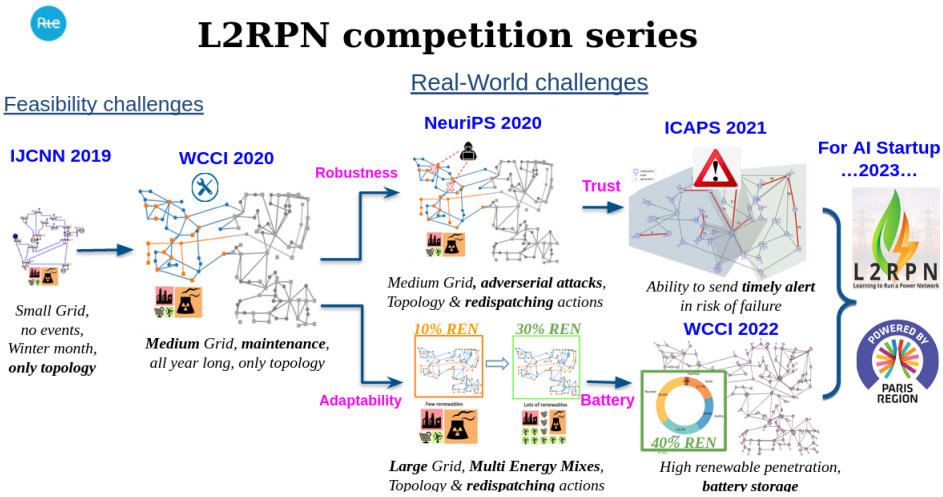

Figure 7: Series of L2RPN competitions since 2022. Each one scaling up the problem in terms of size and realism.

edition, we introduce more realistic scenarios proposed by RTE to reach carbon neutrality by 2050, retiring most fossil fuel electricity production, increasing proportions of renewable and nuclear energy, and introducing batteries. Here an artificial agent should show how well it is able to control flows in power lines on the grid to avoid blackouts.

5.2 Problem formalization

From a theoretical point of view, the L2RPN problem can be seen, at first glance, as a Markov Decision Process (MDP) well known in the Reinforcement Learning framework as depicted in Figure 14.

An MDP is defined as a tuple (S, A, p, r), where an agent interacting with an environment observes a state $s_t \ 2 \ S$ and takes an action $a_t \ 2 \ A$ at time step t. From state s_t and taking an action a_t, the agent arrives in a new state s_{t+1} of the environment with probability $p(s_{t+1} j s_t, a_t)$, and receives a reward $r(s_t, a_t, s_{t+1})$ as an instantaneous signal on the "quality" of its action.

The environment considered in L2RPN is "episodic", meaning it lasts a finite number of time steps T_{end}. Note that T_{end} is not necessarily deterministic (i.e. known before the start of "the game"). Indeed, for the L2RPN environment, T_{end} depends on the actions done by the agent but also on random events (e.g. random line disconnections). T_{end} is variable but bounded since it can not exceed a fixed maximum duration T_{max} (equal to one week in the challenge): $1 \quad T_{\text{end}} \quad T_{\text{max}}$.

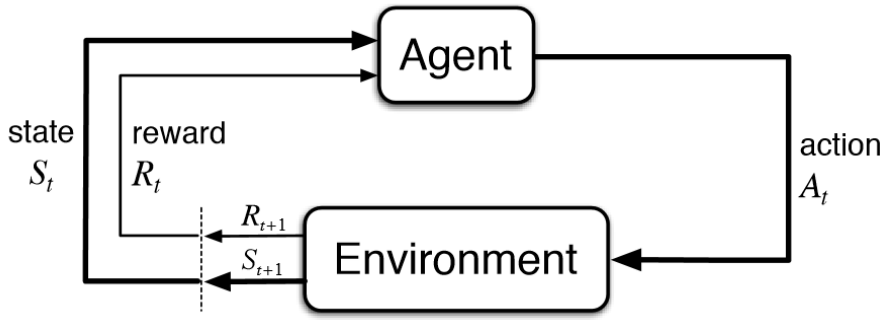

Figure 8: The classical RL framework of an agent interacting over time with a system's environment under an MDP. Note that any kind of agent can be designed and evaluated, such as heuristics or optimization, and it is not restricted to RL agent (Figure from [18])

To act on the environment (i.e. to choose its actions), the agent must have what is called a "policy": A mapping from states to actions $\Pi(s_t) = a_t$. This mapping can be deterministic (a state/action (algorithmic) association) or stochastic (a probability distribution among the actions. The distribution depends on the state).

In the MDP setting, the agent must find a way to maximize the reward it gets from the environment, not only for the current step, but over the whole episode [5] . Its goal is to maximize the total amount of (discounted) reward accumulated over the episode. The MDP problem can hence be stated as "finding the optimal policy" to maximize the (expected) cumulative (discounted) reward over the episode(s).

By construction, MDPs respect the Markov property, which states that all the information necessary to generate s_{t+1} can be found in s_t and a.

In the L2RPN environment as implemented in the grid2Op framework, however, the agent is not given the exact (physical/electric) state s_t of the environment at time t, but just an observation o_t of this state.

In absolute terms, this observation can be considered incomplete (it could be completed, for example, by more electrical data on the power network or by known information about the planned production schedules of the generators) and/or imperfect (limited precision of the sensors, measurement errors, noise, imperfect system model, etc.).

This makes the L2RPN setting not strictly and formally a Markov Decision

[5] It is important to see that the reward received at each moment does not depend only on the action performed at that moment but on the whole action sequence from the beginning. Some "good" actions may not be rewarded until late in the episode. This makes the problem harder than when there is an immediate association between the action and the reward, as it is the case, for example, in the (Machine Learning) classical bandits problems.

Process (MDP) but rather what is called a Partially Observed Markov Decision Process (POMPDP). In practice, however, considering L2RPN as a "simple" MDP is not a bad approximation to find, most often a near-optimal policy. Yet to solve it, one needs to consider several resources to manage constraints associated with it in the form of a credit assignment problem. Available actions or times before disconnection or recovery are examples of budget to deal with. To manage it, an agent has to make a sequence of decisions, with rewards that may only come at some later timestep, such as only when a blackout has eventually occurred.

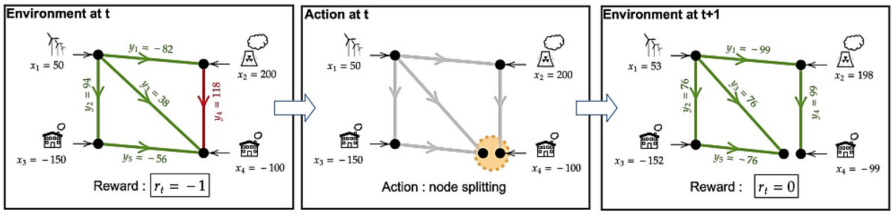

Figure 9: L2RPN quick overview: An agent first observes the power network state at time t with flows in the power lines and injections (productions and consumptions). An overload (red line) occurs on the grid: the agent gets a negative reward from the environment since the grid is at stake. The agent takes a node-splitting remedial (discrete) action at time t. Three node splitting actions would have been possible at this time step. It is easy to see that the number of possible node-splitting actions increases exponentially with the number of switchable elements. After this action, the problem (line overflow) is solved, and the agent gets a better reward.

To continue on the formalization aspects of the problem, we can define an "episode" $\boldsymbol{e}$, successfully managed by an agent up until time $\boldsymbol{T}_{\mathsf{end}}$ (over a scenario of maximum length T_{max}) by the sequence:

$$\boldsymbol{e} = (o_1, a_1, o_2, a_2, \ldots, a_{T_{\text{end}}\ \ 1}, o_{T_{\text{end}}}) \tag{1}$$

where o_t represents the observation at time t and a_t the actions the agent took at time t. In particular, o_1 is the first observation, and $o_{T_{\text{end}}}$ is the last one: either there is a game over (e.g. a blackout) at time T_{end} or the agent reached the maximum time of the scenario: $T_{\mathsf{end}} = T_{\mathsf{max}}$.

At the heart of an MDP (or a more complex POMDP), lies the environment. Most of the time, this environment is implemented through a simulator. Indeed, in L2RPN, we need a simulator that can accurately mimic the behavior of a power system, regarding physical laws and operational rules, over a specified time period referred to as a *scenario*.

To achieve this, the simulator must be provided with data, specifically time-series data describing the electricity (power) injections in the network. These data are referred to as *time series* or *chronics*. Note that this last word derived

from the French *"chroniques"* must be understood in the L2RPN context as just "time-series data". These chronics can sometimes bring high stochasticity, such as with wind power or when considering unexpected line disconnections to be robust too. This makes solving this overall MDP challenging. Figure 9 illustrates the L2RPN MDP problem.

The next section will detail the specific characteristics of the simulation environment (simulator) used, as well as the chronics. Additionally, we will also describe how the competition is organized on an online platform, including the metrics used to rank and evaluate participants.

5.2.1 The simulation environment

Grid2Op The L2RPN competition requires a library/module that can simulate a power system within a reinforcement learning framework. To meet this requirement, RTE has developed Grid2Op [4], a Python module that converts the operational decision-making process in a Markov Decision Process (S, A, P_a, R_a) [2] setting as described in section 5.2. This module discretizes the time in 5-minute time steps (this is representative of the operational process used today in which operators receive new grid snapshots every 5 minutes). For example, a one-day scenario would be divided into 24 $60/5 = 288$ time steps. Given a current state s_t 2 S and an action a_t 2 A, from the agent, Grid2Op can calculate the power flow (the amount of electricity flowing on each power line) at the next time step s_{t+1}. To do this, it will need the time series at time $t + 1$. We detail the generation of these (time-series) data in the next paragraph. Additionally, Grid2Op can use the Gym interface developed by OpenAI [3] to interact with an agent. A set of starter notebooks is provided with the grid2Op (Python) package to simplify the process of developing an agent. The use of these notebooks allows the participants (to L2RPN the challenge) to create efficient agents with limited prior knowledge of power systems, making the competition accessible and geared more towards AI (more specifically Reinforcement Learning) than electrical engineering.

The grid In this year's challenge edition (2023), we use an adapted version of IEEE 118 grid with high renewable penetration [16]. It can be visualized in Figure 10 with productions and consumptions. Its grid topology allows for more than 100 000 unitary actions, the basis for a huge combinatorial action space.

Energy Mix The set of productions allows us to define what is known as *the energy mix of the system*. It is the share of each type of production (nuclear, thermal, wind, solar, etc.) in the total electricity production over a considered period, such as over a year. This year's competition represents the target electrical mix reached by 2035. While today renewables (apart from hydroelectricity) represent 10% of the total produced electricity[6], the L2RPN'2023 challenge will propose to handle the power grid with a higher percentage with up to 30%[7].

[6] Bilan electrique 2021 par RTE

[7] Analyse des scenarios RTE - Commission particuliere du debat public

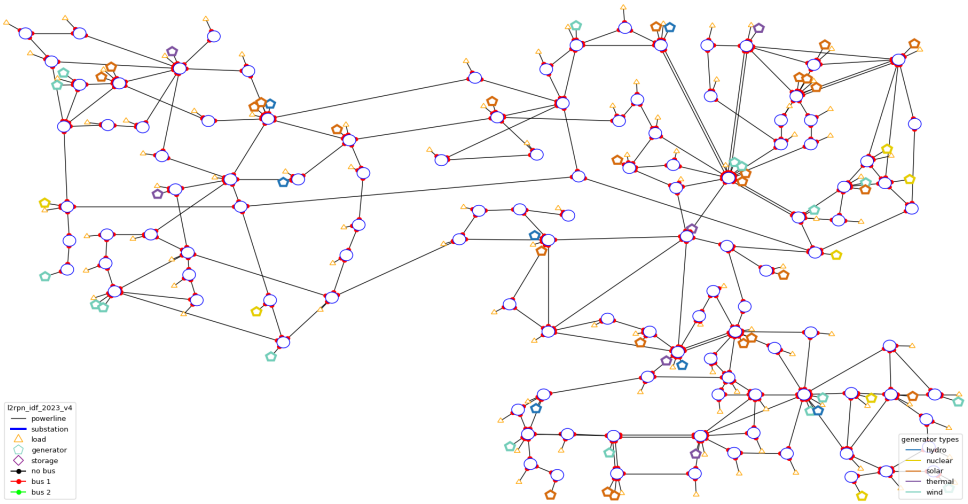

Figure 10: IEEE 118 grid illustrated with the different generator locations with types (solar, wind, hydro,thermal,nuclear
)

Chronics In order to make Grid2Op work and represent the target energy mix, we must have time series describing the electricity injections into the power grid, referred to as *chronics*. These chronics provide the amount of electricity injected into the network by generators, loads, and batteries at each time step. Generators inject a positive amount of electricity while loads inject a negative amount, and batteries can inject either a positive or negative amount depending on whether they are storing or delivering electricity. It is important to note that these injections considered with the energy losses (due to the Joule effect in the grid) must always sum to zero for the power grid to function properly. To generate these chronics, we need a detailed description of the architecture of the power grid: Consumption points (loads) and power generators (type, maximum production, operational constraints, etc.). We also need realistic data about weather conditions (temperature, wind speed, cloudiness, etc.). These data are mostly obtained from past and current RTE studies.

The Chronix2Grid (Python) package, created by RTE, uses these data about the grid to generate time series data, an example of which can be seen in Figure Fig. 12.

The set of productions included in the time series allows us to define what is known as *the energy mix of the system*. It is the share of each type of production (nuclear, thermal, wind, solar, etc.) in the total electricity production over the considered period.

In order to generate the 2023 Challenge edition input time series, we have prioritized the use of renewable energy and set penalties on the use of fossil fuel generators (referred to as *thermal*) in Chronix2Grid. As a result, we were able

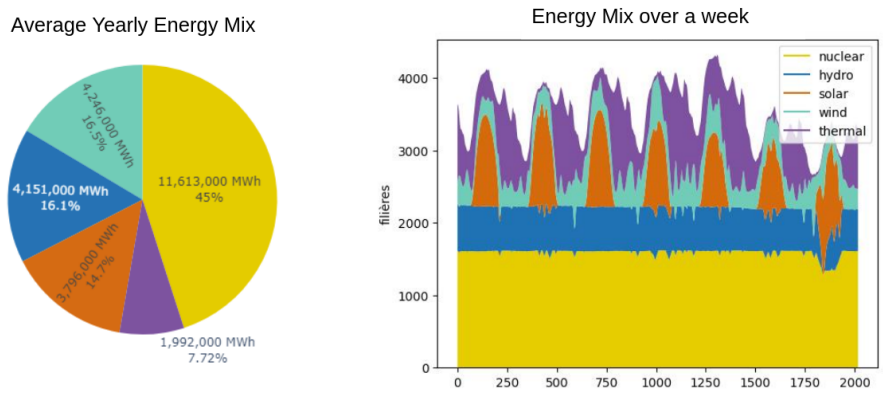

Figure 11: L2RPN 2023 energy mix averaged over a year and along a week. It is representative of expected mixes in France for 2030-2035

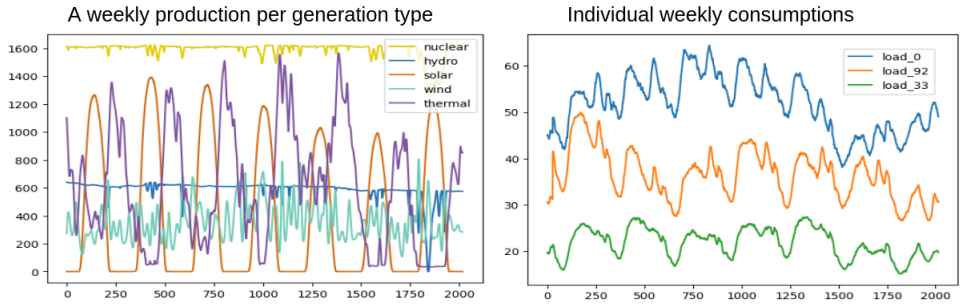

Figure 12: Example of time series representing the energy produced by each type of production at each time step, as well as individual loads. This illustrates their different dynamics.

to create the time series with an almost carbon-free energy mix, with less than 8% of electricity being generated by fossil fuels.

We provide a few dozen of years worth of scenarios for participants to train their agents on. Additionally, through the chronix2Grid package, we allow them to generate more scenarios with the same specifications (energy mix and network parameters).

5.2.2 Setting recapitulation

We summarize the setting of the problem to be solved illustrated in Figure 13:

- **Environment**. The environment is episodic, running at a 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

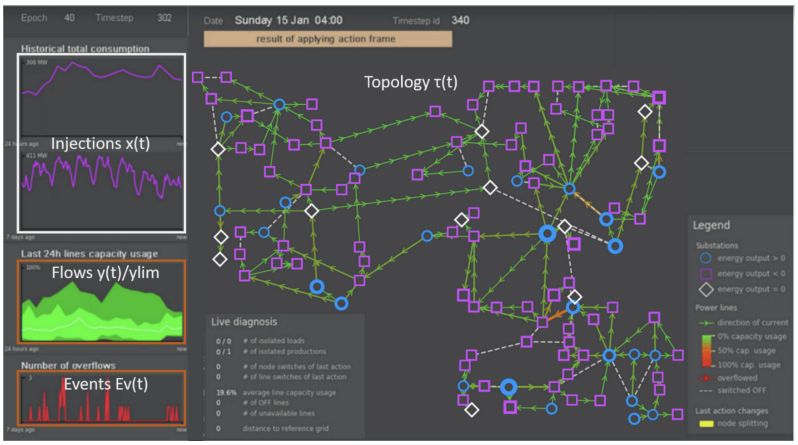

Figure 13: Illustration of L2RPN Environment: productions and loads $x(t)$, events $Ev(t)$ and a grid topology $\tau(t)$ induces flows $y(t)$ in every power line that the agent needs to manage.

graph. The industry standard synthetic IEEE 118 network is used for this competition. Realistic production and consumption scenarios are generated for the network (input time series).

- **Observation space**. Complete state of the power network: all information over power nodes (electricity produced and consumed) and flows of each power line. A more detailed description is given in Section A.3.

- **Action space**. Four types of actions are allowed to intervene on the grid:

  1. Line switching actions: connection/disconnection.
  2. Topology changes (node splitting). (2)
  3. Power production changes/curtailment.
  4. Storage actions (storage or delivery from batteries).

  It is worth noticing that the action space is very large when compared with some well-known classical problems from the Reinforcement Learning literature. Indeed it contains tens of thousands discrete actions (Line switching actions and topology changes) along with a high dimensional (almost 80 dimensions) continuous action space (production changes and storage actions). A more detailed description of the action space is given in Section A.4.

- **Operational Rules** It is also important to know that to keep close to the reality of power grid operations, actions in the grid2Op environment must meet several conditions before being executed. As an example: An agent can not act on the same line twice before a cooldown time. The same

condition applies to topological changes: The agent must wait a certain time before acting on the same substation. There are also limitations associated with changes on the production levels (ramp-up / ramp-down) of the generators. All these limitations and constraints are detailed in the grid2Op documentation. Finally, an other rule in the environment determines how much time a line can be overloaded before automatic line disconnection by protections.

- **Reward**. The reward is constitutive of every MDP and is probably more important than usually thought. It is important to set it carefully because it can have an important impact on the ability of an agent to learn efficiently. This is why we give the participants the ability to design their own reward function.

  However, the grid2Op framework comes with a set of predefined reward functions. Some of these grid2Op rewards are also used to compute the leaderboard metric. This is detailed in Section 6.4.1.

- **"Game over" condition**. What we call a "game over" refers to the feared event where the network is unable to fulfill its main mission: meeting the power demand from the production. The agents must obviously avoid it. Note that the most severe (and common) way to end with a game over occurs when some lines overload (i.e., the electric current going through them exceeds their physical capacity), leading to their brutal disconnection from the network and putting an extra stress on the remaining lines, which in turn leads to more and more overflows (cascading failures) ending with what is called a blackout. The whole process can be very fast (less than a time step), so the agents must be very careful about line overflows.

  In the grid2Op environment, a game-over is triggered if the total electricity demand is not met anymore (Significant production/consumption imbalance). This is taken into account in the leaderboard metric as a (penalizing) blackout cost.

- **Expected and unexpected events**. To get close to a realistic situation and to make the competition more challenging, the simulation of the power grid includes some additional "expected" and "unexpected events". The expected events are related to the planned maintenance of the power grid. From time to time, some lines are switched off for some (fixed) duration to allow their maintenance in safe conditions. "Unexpected events" are related to equipment failures on the network. These failures are simulated in the grid2Op framework by having a special agent called "the opponent agent" with the mission of disconnecting some lines at random in the network. At random but not "completely at random" as it is constrained to keep a reasonable level of aggressiveness to avoid turning the (very serious) control of the power grid problem into some kind of unrealistic "arcade game". Not completely at random also, because the opponent

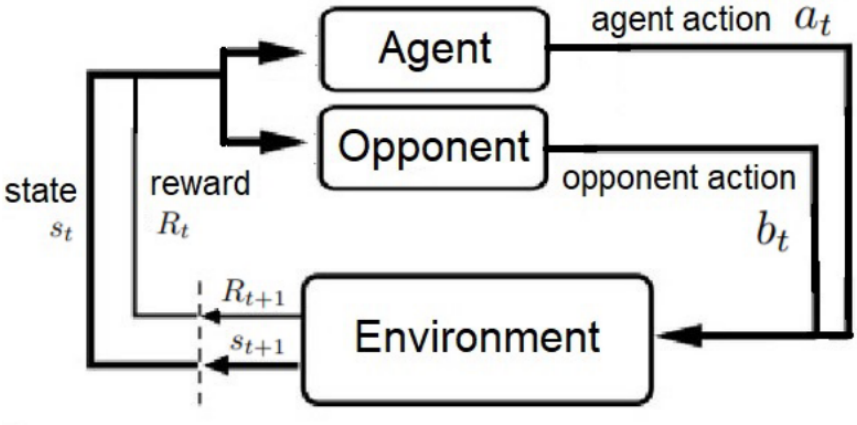

Figure 14: The RL loop with the opponent (Figure from [14]). Note that in the 2023 competition, the opponent does not actually use the reward signal as the (general) figure seems to suggest. It is still a "simple opponent" that uses a predefined static strategy. In particular, it does not learn.

chooses preferably the most electrically loaded lines to disconnect them. As the state of the electrical network depends on the past actions of the competing agent, we see that there is a causal link between the actions of the agent and those of the opponent. Formally, we can see the L2RPN interaction loop extended as in figure 14

- **Forecasts and simulation**. In real network management, operators often use simulations to test their future actions, based on consumption and/or production forecasts on a particular point of the network. It is therefore coherent to allow an (artificial) agent to have such forecasts in the grid2Op environment in order to find and test its actions.

  Until recently, the gri2dOp environment only provided consumption and production forecasts for the next time step, 5 minutes ahead. These forecasts are in fact implemented in the "simulate" methods [8].which allow to simulate an action on the next time step based on forecasts for this moment.

  In the new version of grid2Op, the environment will provide forecasts on a longer horizon of 60 minutes.

  Of course, these forecasts are not perfect. They are subject to errors, calibrated to be comparable to the forecasts used today in the management of the real power system.

[8] See the "simulate" method of the "Observation" class as well as the more sophisticated \Simulator" class in the Grid2Op documentation

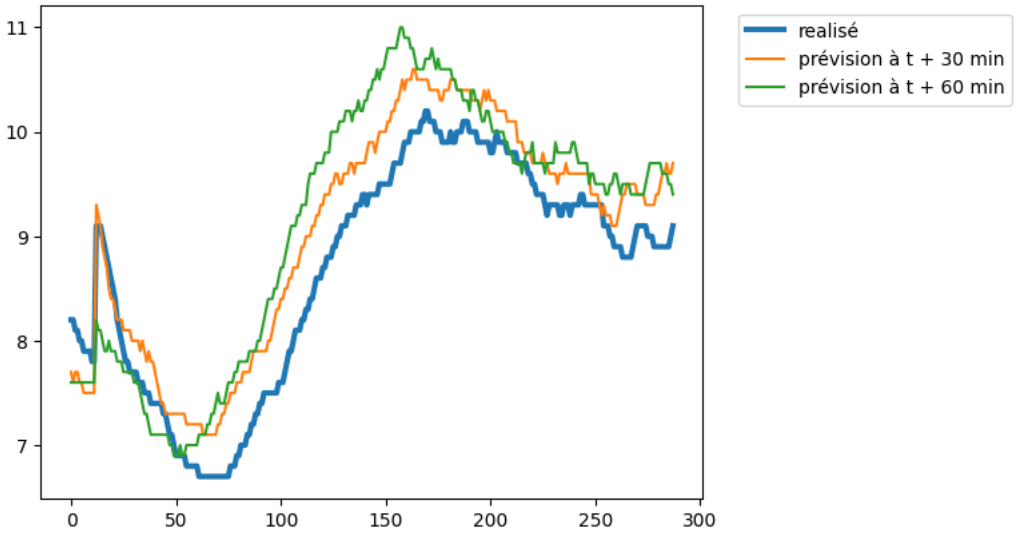

Figure 15: Example of load forecasts at different time-horizons: the real load and the one that were forecasted with 30 and 60 minutes time-horizon. We observe than 60-minute forecast is less accurate than 30-minute forecast on average as expected.

5.3 Lesson learned from previous competition

Over the environment design, competition phase, and feedback iterations and improvements over the years, we can highlight some interesting insights and lessons to build upon.

ML and RL as a promising technology for near real-time sequential and combinatorial topology control The initial L2RPN competition [11] demonstrated the potential of ML and RL to help control the topology of the grid (a high-dimension discrete and combinatorial action space) [7] while taking into account operational constraints. At this stage, on the smallest IEEE14 grid, expert rules could still achieve decent performance, ranking 5th. This changed when scaling up the problem for NeurIPS competition [10], where except for an advanced adaptive expert system [8], no hard-coded rules could achieve interesting performance. ML and RL also showed an advantage in terms of inference time compared to optimization approaches which is an important consideration for near real-time decisions. The winning team from Baidu showed an interesting Deep RL with evolutionary training solution [20] that won by some margin on the robustness track. Yet, those more advanced agents would still fail on one-third of week-long test scenarios. This failure rate was a bit improved in the latest competition by a solution transposing AlphaZero approach to grid topology control [5], but not nearly close to the level of reliability required to

deploy autonomous agents on the grid. They will rather be used under human supervision.

Nevertheless, one striking feature was their ability to sometimes meaningfully combine sequences of 5 or more topological actions to survive very difficult situations when other agents and perhaps even a human will fail. Such combinatorial action depth is actually beyond what humans are able to explore today and can already be viewed as a unique achievement.

Reducing the action space Except for the Adaptive Expert agent, all other agents (learned, optimized, or heuristics) have considered as a first step a reduction of the topology action space below 1000 unitary actions. A learning agent that does not disregard some actions before training remains to be seen. Perhaps disregarded actions would have proven to be useful and helped better manage the grid in some situations.

Mixing discrete and continuous actions Properly combining those actions is important. On the one hand, discrete topological actions are often preferred since they are cheap and carbon-free. The Alphazero-like agent demonstrated that it could avoid up to 90% of C02-emitting redispatching remedial actions. On the other hand, they are not always enough to solve difficult issues. In that case, continuous actions should be combined. The latest agents have mixed topological actions with more continuous ones (redispatching, curtailment, batteries) but as separate modules that are merged in an ad hoc fashion. Control with continuous actions has so far relied on optimization formulation considering only the current timestep (known as optimal power flow). A learning module for continuous actions remains to be seen, not to mention an agent learning jointly to control those two kinds of actions in a more optimal fashion.

The opponent - the main trigger of blackout In real-life, dealing with unexpected line disconnections that makes the grid weaker, decreasing its overall electricity transfer capacity, is one of the most challenging. In our environment, most of the agent failures are also observed after opponent attacks [14] that corresponds to unexpected line disconnections. Being robust to such events is one of the main challenges for agents in operating a power grid. Those agents hence need to learn effective strategies. To do so, most advanced agents often use preventive actions ahead of attacks to mitigate the risk of the most difficult ones.

Using simulation for safe decisions One particularity of our problem setting is that the agents can use a simulator at inference time. However, limiting the use of simulation is preferable to make faster decisions in a given time budget. Having a fast inference is where ML actually makes a difference. Yet all agents have so far used some residual simulation to robustly validate their choice of action. Most often, a model infers the most promising actions, but the final choice is then only done based on the results of their simulations for the next

time step. This seems a reasonable approach that human operator actually uses: when they are about to run an action on the grid, they simulate it one last time, as the cost of doing a wrong action can be high.

Bigger computing budget for training wins In previous competitions, the participants used their own infrastructure for training their agents. The winning agents often required significant computing resources to be trained. That is over 100 CPUs simulating episodes in parallel to learn from, with a training run of at least one day. Having a bigger computing budget during training most probably made a difference. This year, we would like to make fairer comparisons through the sim2Realistic track (see subsection 6.2. In particular, in this track, it is expected that participants train an agent with the same computing budget we give them.

Introducing the assistant feature In high-security environments such as the power grid, agents are not expected to be ever deployed completely autonomously, i.e., to operate the grid without human supervision. But they are still expected to be very useful if used within an assistant to human operators. To facilitate such an integration, such agents need to be trusted. It is hence important that the agents estimate their degree of confidence when they make action suggestions. To that end, we introduced an additional task: predict confidence. This will be evaluated in the Assistant track (see section subsection 6.3)

This year assistant feature is an improvement and partly a reformulation of the first iteration on this idea run for the ICAPS competition [9]. A drawback of the ICAPS approach, which required the agent to issue an alarm 30 minutes prior to an anticipated blackout, was that it resembled predicting the timing of an attack - a task that is inherently unpredictable.

In this year's challenge, we rather formulate it as "If this attack happens, are you confident in your ability to manage the grid in the following steps ?". These alarms are also now considered at the granularity of lines as in operational applications today and not at the level of areas.

Another proposed assistant feature by another team is the ability to make several recommendations and express preferences over the kind of actions the assistant should highlight (a weighting factor between topological and redispatching actions) [6]. This is of interest but will not be directly evaluated in this competition.

Undesirable behaviors Over competitions, we have seen some undesirable behaviors from various agents that would be regarded as unacceptable for human operators. Some agents sometimes oscillate between configurations (or actions), for some period of time. This can be seen as unstable and risky, as in managing critical systems the least actions to stable reference configurations are preferred.

When in difficult situations, some agents also just desperately run topological actions, which could be interpreted as a sequence but which just happens to be a very reactive or greedy behavior. Running through topological actions

with such high combinatorial depth is not desired except if really necessary. Operators prefer more straightforward and simpler solutions to stay around more well-known configurations. The Grid2viz and Gird2Bench packages (see subsubsection 6.5.3) are tools to study more in-depth such behaviors.

Limiting those behaviors would hence make the respective agents more trustworthy. Would it be through more constraints in the environment or better cost functions? This remains an open question.

We also a posteriori discovered pitfalls in competition design for two competitions at WCCI 2020 and WCCI 2022. At WCCI 2020, the winning approach learned this strategy, which allowed to disconnect and reconnect lines more frequently, using this to circulate the overload between power lines until the situation is less tense. This was not really solving the problem the way an operator would expect. This came from a limitation in the environment at that time, that was then improved. In last year's WCCI 2022 competition, one key factor to be among the winning teams was to curtail as much renewable energy as possible. This is not a desired behavior to tackle Climate Change in the end. It was not penalized enough, and we made a better cost function this year in that regard.

Participant feedback appreciated In this year's competition, we encourage participants to provide an analysis of their agent behavior, feedback on environment limitations, and suggestions on possible improvements, much like a collaboration with the organizers, which will matter once the project of the winning team is launched. This will be considered in the evaluation (see subsubsection 6.4.5).

6 Description of the new competition setting

6.1 A competition in two tracks

For this competition, our aim is to test various aspects of the proposed methodology. Each participant will be required to submit entries for two distinct tracks:

- The "Sim2Real" track, where we will evaluate the adaptability of the agent to function in an environment different from the one it was extensively trained on.
- The "Assistant" track, where we will assess the agent's ability to collaborate efficiently with a human operator.

Both tracks are described in more detail in the following subsections.

6.2 Sim2Real track

In all the previous "L2RPN" competition series, we used a simulator to model the power grid and to carry out both the training and the evaluation of the

agents. This is justified because the agents cannot be trained in the "real world", for obvious security reasons. However, when deployed in practice, an agent will have a simulator available for training and for probing the effect of proposed actions at inference time (by using the *observation.simulate(action)*), but the effect of its action will have consequences in the real world. No matter how accurate a simulator is, it will never perfectly emulate the real world. Hence, our previous evaluation setting biased results favorably.

To remedy this problem to some extent, we make use of two simulators: a more detailed one emulating the real-world that we call "real-world emulator" and a simpler one representing a "simulation tool" . The "simulation tool" is made available to the agents for training and inference. The "real-world emulator" is used by the organizers only for evaluation purposes. This is implemented in practice by leveraging different fidelity of our power grid simulator *grid2op* that will represent more or less complex dynamics. It is simple to toggle the simulator behavior for the "simulation tool" with the *obs.simulate* function, a feature also made available to the agents.

We expect, with this new more realistic setting, that the performance of the agent will degrade if only trained with a "simulation tool" with no more consideration of the real-world. De facto, they will have to perform some kind of transfer learning to be ready to adjust to distribution shifts. In the real setting, agents would be able to retrain and fine-tune their agents only based on pre-recorded real-world historical data. As no such historical data is available on such synthetic grid, we make our competition setting close but somewhat different from it, by allowing agents to generate their own historical data on a limited number of scenarios while interacting with the "real-world emulator". This is why this track is named "Sim2Real": agents can be retrained before being evaluated in this more realistic setting.

To summarize, the goal of this track is to assess whether agents can be generic enough to be used on the real grid. To do that, participants will submit agents that will be tested by the organizers on an environment that uses a more realistic simulator than the one available to the agent at training and inference time. Before being evaluated, the agent will have the possibility to be fine-tuned on the environment used for evaluation by the organizers. We emphasize that from the "grid2op" point the "training" power grid and the "test" power grid will have exactly the same properties: same elements at the same location, same number of actions, same operational constraints etc. Only the physical parameters and the underlying simulator will be different.

6.3 Assistant track

The assistant track will require the agent to send warnings by anticipation to a fictitious human operator. These warnings should be raised in case the agent is not confident about its capabilities to run the network safely for the next 60 minutes, if a contingency event occurs. The list of considered contingencies is the set of attackable lines in the competition. At each timestep, the agent can decide to send a warning, or not, for each considered contingency. A desired

behavior is that an agent should not send too many warnings to keep the human operator concentrated on his or her job, and that a warning should be persistent through time (not switching on and off from a time step to the next). To be consistent with this behavior, the agent is given a limited "attention budget" which will limit the number of allowed warnings. More precisely, the attention budget decreases, when the agent warns the human. This element is further detailed in subsection A.6.

This assistant feature will be evaluated in the next 60 minutes an unexpected line disconnection has occurred: either a blackout happen or not, hence the confidence of the agent was right or not regarding that contingency. In this new competition, 60-minute forecasts are made available and could be used for this assessment. See Figure 16 for an example of this mechanism explained. You can further refer to subsubsection 6.4.4 for more details about the evaluation score.

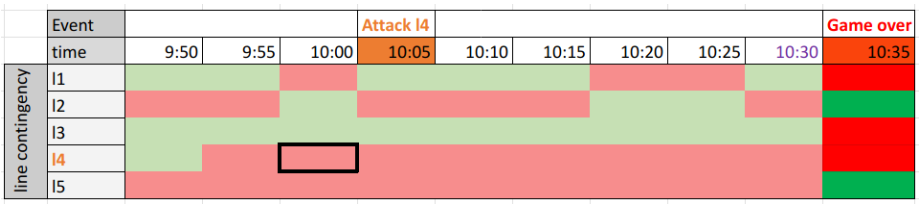

Figure 16: Example of agent warnings in red over time for 5 considered line contingencies (among attackable lines). One attack happened on l4 at 10:05 and the agent had a warning raised just before indicating it would get into trouble if that was the case. Indeed at 10:35, less than 60 minutes later, a blackout occurs. The agent was right about its low confidence and gets a positive evaluation on this example. If it managed to continue beyond 11am, more than one hour after the attack, it should have been confident. As it was declared not confident, it would have got a negative evaluation

6.4 Evaluation

6.4.1 3-dimensional score for quantitative participant evaluation

To rank the participants, a **score function** is applied to evaluate the agent's performance through a numerical score.
The score for this competition will be exhibited along three dimensions that are explained in the further paragraph:

- **Operation score**: it is based on the cost of operations of a power grid that includes the cost of a blackout, the cost of Energy losses on the grid, and the cost of actions. It ranges between $[\ \ 100, 100]$.
- **Low-carbon score**: it is based on the amount of renewable energy curtailed. The less renewable energy curtailed the more carbon efficient the

grid operation is. It ranges in $[\ \ 100, 100]$ with 0 meaning "renewable energy sources have not been used at all (they have been entirely curtailed) and 100 that every possible MW of renewable energy is used.

- **Assistant score**: it is based on the number of times an agent is right about its confidence ahead of time to handle some contingencies, specifically line disconnection events. It also ranges in $[\ \ 100, 100]$.

Each of these scores are detailed in below sections.

The overall score will be a weighted sum of these standardized scores, such as

$$Score = 0.5 \quad \boldsymbol{Score}_{\textbf{Operations}} + 0.2 \quad \boldsymbol{Score}_{\textbf{Low-carbon}} + 0.3 \quad \boldsymbol{Score}_{\textbf{Assistant}} \tag{3}$$

In any case, it should always be beneficial to complete a scenario rather than falling into a blackout.

6.4.2 Operation score

As power grids are more and more modeled as live market exchanges, almost all of the grid's operational characteristics can be converted into a monetary cost. Therefore, the score will reflect the realistic operational costs of a power grid, which grounds the algorithmic performance of proposed agents in a very real-world quantity: money.

- **Energy Losses Cost**: determined by multiplying the total electricity lost due to the Joule effect (in Mwh) by the market price of electricity (in EUR/MWh).
- **Flexibility Cost**: the sum of expenses incurred by the agent's actions. Changes in electricity production (*e.g.* curtailment or redispatching) have a cost that varies based on the energy market, while using storage units has a fixed cost per MWh.
- **Blackout Cost**: in case the agent fails to manage the power network until the end of the scenario. This cost is calculated by multiplying the remaining electricity to be supplied by the market price of electricity.

It should be noted that the cost of a blackout, as expected, is significantly higher than the other two costs. This means that an agent who successfully completes a scenario will almost always have a higher score than one who does not, even if their actions are less expensive.

Yet energy losses cost RTE 500 million €/year, so **a gain of 20% would already save 100 million €/year**. And if no flexibility is identified or integrated on the grid, operational costs related to redispatching can dramatically increase due to renewable energy sources as was the case recently in Germany with **an avoidable 1 billion €/year increase**[9] illustrated on Figure 17. For

[9] German power system operational cost https://allemagne-energies.com/2018/06/19/allemagne-14-milliards-deuros-pour-stabiliser-le-reseau-electrique-en-2017/

more details about the equations of these underlying costs, please refer to section B.5.

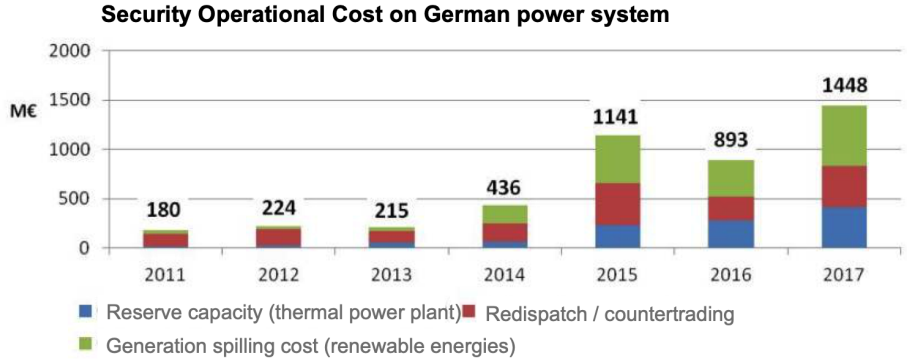

Figure 17: **Undesirable 1 billion/year operational cost sharp increase in Germany** in recent years after quick installations of renewables without developing new flexibilities

Final cost of operations normalized Now we can define our cost $\boldsymbol{c}$ for an episode:

$$c_{\text{operations}}(\boldsymbol{e}) = \sum_{t=1}^{T_e} (c_{\text{loss}}(t) + c_{\text{exibility}}(t)) + \sum_{t=T_e}^{T_{max}} c_{\text{blackout}}(t) \tag{4}$$

The score of an episode is then a function of this cost, which is designed to be maximized (-cost) and is scaled to fall within the range of $[\ \ 100, 100]$. The score function is also normalized to have a score of $\ \ 100$ for an agent failing directly from the start, of 0 for the reference `doNothing` agent, the agent that takes no action at all and often ends up failing. 80 being the score of an agent succeeding at the scenario with no improvement in energy losses and only using cheap actions (*ie* no action on redispatching, storage units or curtailment), and 100 the score for an additional 20% gain in energy losses (while still not using any costly actions).

6.4.3 Low-carbon score

The Energy Transition to meet the net-zero carbon emissions objective is primarily driven by the proper integration and use of new renewable energies. This score is here to reflect the ability of an agent over a scenario to use available renewable energies at their full potential. This means avoiding as much as possible curtailement $\boldsymbol{\mathcal{E}_{curtail}(t)}$ of renewable energy $\boldsymbol{\mathcal{E}_{renew}(t)}$, which would imply redispatching on carbon-emitting power plants, hence an increase in carbon-emissions.

$$Rate\_E_{curtail}(\boldsymbol{e}) = \sum_{t=1}^{T_e} E_{curtail}(t) / \sum_{t=1}^{T_e} E_{renew}(t) \tag{5}$$

The score of an episode is then a function of this rate, which is designed to be maximized (-rate) and is scaled to fall within the range of [100, 100]. 100 being the score for no curtailment in the episode, 0 the score for a 20% rate and 100 for a 50% rate.

6.4.4 Assistant score

This score evaluates how confident an agent was in its actions for handling unforeseen line l disconnection events prior to occurring ($confidence(l, t \quad 1)$ True of False), for the next 60 *minutes* (12 *timesteps*) time horizon. Much like a green or red indicator per considered contingencies, analogous to a real application depicted in section 4. When such an event $Ev_l(t)$ occurs for a given line l at time t, the evaluation will look over the duration of the considered time horizon to see how well such an event was handled. To make it more tractable in this competition, we only consider line events among the list of n_l attackable lines, and not all powerlines.

If the event was well handled and ($confidence(l, t \quad 1)$ was True), then the agent gets 1 point. Otherwise, it gets a small penalty of -1 point.

If a blackout occurs, the evaluation considers the earliest disconnection event in the considered time horizon prior to the blackout. The agent gets 2 points if it was right about its confidence for that earliest event. Otherwise, it gets a high penalty -10.

We hence have a trust score for each event occurring at some time t_k over some line l:

$$c_{trust}(\boldsymbol{Ev(t_k)}) = \begin{cases} 1, & \text{if } confidence(t_k \quad 1, l) \text{ and } no\_blackout([t_k, t_k + 12]) \\ 2, & \text{if not } confidence(t_k \quad 1, l) \text{ and } blackout([t_k, t_k + 12]) \\ 1, & \text{if not } confidence(t_k \quad 1, l) \text{ and } no\_blackout([t_k, t_k + 12]) \\ 10, & \text{if } confidence(t_k \quad 1, l) \text{ and } blackout([t_k, t_k + 12]) \end{cases} \tag{6}$$

A desired behavior is that an agent becomes confident most of the time in its ability to solve situations and limits its number of warnings towards the operator to help him focus its attention. This can be assessed by computing the proportion of cumulated "green" indicators over the episode:

$$confidence\_rate(\boldsymbol{e}) = \frac{1}{T_e \quad n_l} \sum_{t=1}^{T_e} \sum_{l} confidence(t, l) \tag{7}$$

This metric will not be directly considered in the score, but will be provided for feedback. It is nevertheless constrained in the environment through the attention budget mechanism.

Considering the occurrence of n such events in the episode e, the cumulated score is:

$$c_{trust}(\boldsymbol{e}) = \sum_{k=1}^{n} c_{trust}(Ev(t_k)) \tag{8}$$

The normalized cumulated score of an episode is 0 on average for an agent running until the end of a scenario but with random confidence. It is -100 for an agent encountering a blackout without any positive confidence points. It is 100 for an agent succeeding at the scenario and always confident about its ability to deal with the attacks that happened during the scenario.

6.4.5 Other evaluation

For the choice of the Winner, the members of the Jury will assess the ranking of the agents developed by the Selected Candidates and visible on the Challenge platform, and the value of the answer given in the scientific file of the Selected Candidates. The following criteria will be taken into account in particular, listed in no order of importance:

- Good understanding of the problem;
- Frugality and simplicity of learning with an explanation of the process;
- Agent behavior consistency analysis for human assistant use;
- Feedback and suggestion on environment design pitfalls or limitations
- Prospects in terms of economic development and job creation.
- Ability to open-source solution modules

6.5 Competition organization and materials

6.5.1 Starting Kit

To facilitate participation and reduce the entry cost, we provide a **starting kit**: a set of tools and tutorials to help participants getting started. The starting kit is available on Competition SK.

It contains notebooks summarizing the problem and the competition setting, several baseline examples that can be directly submitted, code to locally verify if a submission is valid or not before submitting it on CodaLab and documentation. The sample submissions (the code of a baseline agent) use the competition dedicated API. Sample scenarios (time series of productions and loads) are also provided. They are generated with the same criteria as those used to test the agents on CodaLab, but they are obviously different as the latter are kept undisclosed to the participants.

This is useful because it helps participants understand the problem they are trying to solve and the context in which it occurs. It also provides an overview

of the competition rules and criteria for success, which can help participants tailor their solutions accordingly.

Regarding the simulated environment, the Python library used is Grid2Op [4]. Here is an example of the most basic code, for those familiar with OpenAI Gym, in order to overview the package:

```
import grid2op
# create an environment
env_name = "rte_case14_realistic" # choice of environment here
env = grid2op.make(env_name)

# create an agent
from grid2op.Agent import RandomAgent
my_agent = RandomAgent(env.action_space)

# proceed as you would any open ai gym loop
nb_episode = 10
for _ in range(nb_episode):
    # you perform in this case 10 different episodes
    obs = env.reset()
    reward = env.reward_range[0]
    done = False
    while not done:
        # here you loop on the time steps:
        # at each step your agent receive an observation,
        # takes an action,
        # and the environment computes the next observation.
        act = my_agent.act(obs, reward, done)
        obs, reward, done, info = env.step(act)
```

The training loop can be simplified using a runner, as shown in the following piece of code:

```
import grid2op
from grid2op.Runner import Runner
from grid2op.Agent import RandomAgent
env = grid2op.make()
nb_episode = 10
runner = Runner(**env.get_params_for_runner(), agentClass=RandomAgent)
runner.run(nb_episode=nb_episode)
```

Learn more in Grid2Op documentation[10].

[10] https://grid2op.readthedocs.io/en/latest/

6.5.2 Hosting on CodaLab

The L2RPN'2023 competition is implemented on the CodaLab Competitions platform [15], enabling code submission, detailed outputs and providing a starting kit.

Participants are required to submit their agent on CodaLab so it can be trained and tested on the platform's compute workers. Submitted agents are blind-tested on the platform with new scenarios not known to the participants. These new scenarios are representative of the different problems encountered by power network operators.

The competition is divided into **three phases**:

- (0) Warm-up phase: Participants have the opportunity to test the starting kit, request modifications to the provided computational resources and packages,
- (1) Development phase: The computational resources and packages are fixed, and participants receive feedback on their submissions through a leaderboard. This phase allows the 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 can develop and improve their model iteratively and regularly.
- (2) Final phase: We re-evaluate only the last valid submission of each participant on a new undisclosed dataset: the test dataset. This final evaluation determines the final ranking. Note that the two undisclosed evaluation datasets (validation and test) are specifically built by the competition organizers to be representative of the diverse problems faced by network operators, such as overflows due to high load, or high renewable generation. The validation and test datasets are drawn from the same statistical distributions.

6.5.3 Other Available materials - GridAlive

Regarding Grid2op, you can go through the getting started notebooks to get a large tour of the framework features. For more information, also see its comprehensive documentation [11]

You can refer to gridAlive[12] platform as depicted in Figure 18 to find out about these other relevant resources and materials in the Grid2op ecosystem. You will find tools for data generation, agent baselines, agent analysis and episode (re)play as well as for faster simulation.

[11] See Grid2op documentation:https://grid2op.readthedocs.io/en/latest/

[12] github GridAlive: https://github.com/rte-france/gridAlive

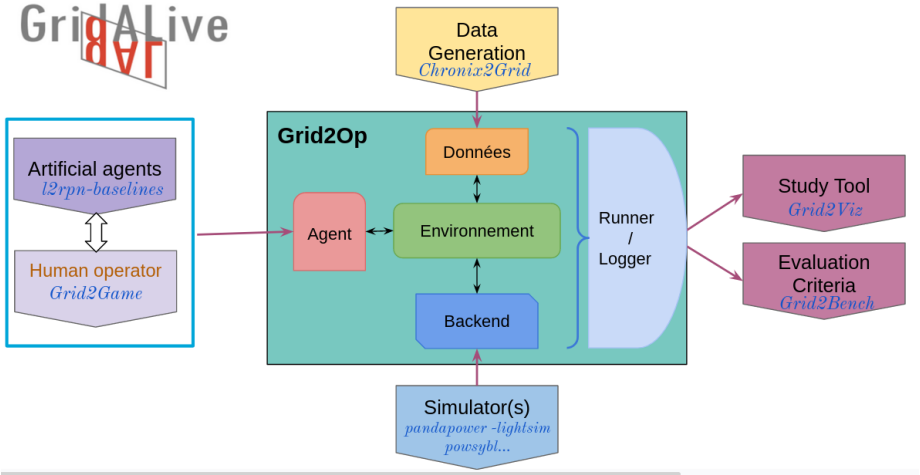

Figure 18: GridAlive - Grid2op ecosystem with packages and materials available

Appendices

A Detailed description of data and simulation environment

A.1 A Power Grid

This section focus on the description of the specificities of the electrical network designs for the L2RPN competition.

Generators On a network, power is provided from multiple technologies using different fuels which are all referred to as generators. They can be considered as sources in the power network. Traditionally power was generated by large thermal units burning fossil fuels such as coal, oil and gas. In recent years due to the shifts in policy to decarbonize society and in the liberalization of electricity markets, generation sources have shifted to renewable, unpredictable, weather-based sources such as wind and solar. These sources are often installed in geographically diverse and less populated areas and produce power far away from load centres. Hydro and nuclear power stations, while not new, are carbon-free and are also located relatively far away from load centres. The network needs to be planned and operated in different ways, to incorporate geographically disperse, variable generation sources and ensure that power is efficiently transmitted to the load centres at all times.

Transmission The transmission network is made up of transmission equipment, overhead lines, underground cables, and substations that contain the connections to generation sources and transformers. Electricity is transmitted either using DC (Direct Current) transmission or AC (Alternative Current) transmission system. Both AC and DC allow transmission at high voltage (thus reducing current), but the main reason for using AC in power networks is that it allows the raising and lowering of voltages using power transformers. Being able to increase the voltage allows us to transmit electricity greater distances due to the lower resistive heating losses (further detailed in subsection B.1). DC is a simpler linear system than AC, which is more difficult as it introduces non-linearities based on sinusoidal aspects of voltage and current generation and three-phase transmission. The RL challenge will run on an AC powerflow, but understanding DC powerflow is a good starting point to understanding power network control.

Consumptions The goal of power system is to conduct electricty to substations, which then require transformers to lower the voltage to a level suitable for homes and businesses to use. The transmission voltage ranges from 100,000 to 760,000 volts, while the voltage in homes is 220 Volts in Europe and 120 Volts in North America. The power network operator's responsibility usually ends once the power is delivered to the step-down transformer.

A.2 Line Outages

Across time series, events can happen such as maintenance operations and unplanned lines disconnection. In order to maintain the network in a safe state and keep on delivering reliable electricity everywhere even in difficult circumstances, actions must be taken to keep/restore the network safety.

Maintenance - Planned outage Maintenance operations are regularly scheduled on the grid. When a powerline is "in maintenance", the powerline is made unavailable. During this time window, it cannot be reconnected by the Agent before the end of this maintenance. These events are planned and information about future maintenance is available in the Observation: time of the next planned maintenance and its duration.

Opponent attacks - Unplanned outage In this serie of challenges, unplanned line disconnections are modeled as an "opponent" who will attack in an adversarial fashion some lines of the grid at different times (similarly to cyber-attacks for instance)[14].

Its role is to simulate failures on the network at particular times. Thus, the proposed agent must overcome these adversarial attacks and keep operating the grid safely. At test time, the agent will eventually be tested against that opponent on hidden new scenarios not presented in the training set, in order to assess the robustness of your agent.

The opponent is designed so as to be as unpredictable as possible, since we do not want the agents to learn and predict the behaviour of the opponent and adapt specifically to it. Attack times are also random, drawn according to an exponential distribution (geometric distribution in discrete time) calibrated to have roughly one attack per day on average but not always exactly one per day as before. However this year will introduce opponents spread over 3 areas of the IEEE 118 grid, called a multi-area opponent. Hence the number of attacks will triple. The durations of the attacks are also changing following an exponential distribution with a within a duration constraint of 2 to 8 hours. See [9] for more details on the opponent.

In order to reflect the idea that the most electrically loaded lines are generally the most prone to failures, we have weighted the probability for a line of being the object of the current attack by the load factor of the line. In this year challenge multiple attacks are possible with a maximum of 3 simultaneous attacks, 3 being the number of opponent areas.

It is important to note that for fairness the attack times and durations are the same for everyone in the evaluation scenarios (even if these times and durations are unknown to the participants), but not necessarily attacks on the same lines.

A.3 State space

At every step, an agent can observe the complete state of the power network. It includes all information over power nodes (electricity produced and consumed), flows of each power lines, and more. After each action, the simulator computes the next state of the environment and the agent creates a new observation. It is described in the Grid2Op environment by an *Observations* with the following information, as described in Table 1

An exhaustive description of observations is provided in Grid2Op documentation.

Future timesteps are associated with forecasts with similar attributes to the observations, at intervals of 5 minutes and for a horizon of one hour.

A.4 Action space

There exists five families of actions accessible in the Grid2Op environment[13]:

topological action : topology can be changed by switching on and off power lines or reconfiguring the busbar connection within substations. The total number of topological configuration is exponential with the number of the lines in the grid. (discrete)

storage action : defines the setpoint for charging and discharging each storage units such as batteries. (continuous)

[13] See Grid2op documentation for more details on the action space

Group	Name and description	Type	Size
Datetime	day	int	1
	month	int	1
	year	int	1
	week_of_the_year week of the year	int	1
	day_of_the_week day of the week	int	1
	hour	int	1
	minute	int	1
	seconds	int	1
Scenario	*current_step* current step in the scenario	int	1
	max_step maximum number of step	int	1
	time_next_maintenance steps to next maintenance	int	*n_line*
	duration_next_maintenance	int	*n_line*
Generator	*gen_p* active power	float	*n_gen*
	gen_q reactive power	float	*n_gen*
	gen_v voltage magnitude	float	*n_gen*
	gen_theta voltage angle	float	*n_gen*
Load	*load_p* active power	float	*n_load*
	load_q reactive power	float	*n_load*
	load_v voltage magnitude	float	*n_load*
	load_theta voltage angle	float	*n_load*
Line origin and extremity	*p_or* and *p_ex* active power	float	*n_line*
	q_or and *q_ex* reactive power	float	*n_line*
	a_or and *a_ex* current flow	float	*n_line*
	v_or and *v_ex* voltage magnitude	float	*n_line*
	theta_or and *theta_ex* voltage angle	float	*n_line*
	rho line capacity	float	*n_line*
	timestep_overflow steps since powerline overflow	int	*n_line*
	time_before_cooldown_line remaining line cooldown steps	int	*n_line*
	time_before_cooldown_sub remaining sub cooldown steps	int	*n_line*
Topology	*topo_vec* bus connection	int	*dim_topo*
	line_status (dis)connection of a line	bool	*n_line*
Curtailment	*curtailment_limit*	int	*n_gen*
	gen_p_before_curtail generator p before curtailment	float	*n_gen*
	curtailment_mw amount curtailed in MW	float	*n_gen*
	curtailment ratio curtailed per generator	float	*n_gen*
	gen_margin_up generator margin up	float	*n_gen*
	gen_margin_down generator margin down	float	*n_gen*
Redispatching	target value	float	*n_gen*
	actual value	float	*n_gen*
Storage details	*storage_charge* state of charge	float	*n_storage*
	storage_power_target power target	float	*n_storage*
	storage_power power	float	*n_storage*
	storage_theta voltage angle	float	*n_storage*
Alarm	*is_alarm_illegal*	bool	1
	time_since_last_alarm	int	1
	last_alarm	int	*dim_alarm*
	attention_budget	int	1

Table 1: Observation description

redispatching action : increase the generator's active setpoint value by either augment or reduce the produced power on each generator. This will be added to the value of the generators. (continuous)

curtailment : allows to give a threshold maximum value to renewable generators. They are defined as ratio of maximal production Pmax. For example, a value of 0.5 limit the production of this generator to 50% of its Pmax. (continuous)

raise line contingency alarm : allows to raise an alarm or not for each line contingency considered in the network. (discrete)

Also see Appendix C for more explanations about some actions and terminology such as substation and busbars.

The targeted assistant is further described in section 4

Note that because a dispatcher is only able to perform a limited number of action at each step, each line and substation is subject to a "cooldown" to limit the number of *topological action* on each element.

A.5 Rewards

Participants are free to design their own reward function. However, the final ranking of the competition is done by computing both the cumulative network operational cost, as well as the assistant cost.

Several pre-defined rewards also are accessible in Grid2Op in order to help the participant.[14].

L2RPN Reward The first standard reward, called *L2RPNReward*, was used in the WCCI competition. It makes the sum of the "squared margin" on each powerline, where the margin is defined, for each powerline as:

$$powerline\ margin = \begin{cases} \frac{thermal\ limit - flow}{thermal\ limit} & \text{if } flow <= thermal\ limit \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

This rewards is then: $\sum_{powerline}(powerline\ margin)^2$.

Assistant Reward While the reward described above focus on maintaining the powergrid in a safe state, other considerations could be taken into account, such as warning when the agent is not able to handle an unexpected future issue on the grid. Towards this objective, the *Assistant reward* was defined: when the environment is in a "game over" state (eg it's the end) then the reward is computed the following way:

1. if the environment has been successfully manage until the end of the time series, then 1.0 is returned

[14] See Grid2Op documentation for the description of other rewards.

2. if no alarm has been raised, then -1.0 is return

This reward was proposed for the past L2RPN ICAPS competition. For this new edition, the alarm score de ned in subsubsection 6.4.4 must be considered and a new reward possibly be de ned.

CombinedReward As the power network is a complex system using only one reward might not be sufficient. Grid2op *CombinedReward* defined as the sum ot the multiple rewards it is made of.

A.6 Assistant Representation

Following the objective of building trust between a digital agent and a human operator, the agent has to communicate when doubting with its ability to operate the network in the future. This element is called the *assistant feature*. Extending the modelling made in a previous challenge [9], the trust objective for this competition is modeled by two main elements : the `raise_line_contingency_alarm` and the `attention_budget`. In this competition, the agent is asked to decide whether or not to trigger an alarm, for each line that could be overloaded in the 1-hour future time window. This alarm is taken into account in the following time steps (for the next 1-hour horizon) using the score described in subsubsection 6.4.4. More precisely, to raise an alarm the agent performs an *action* on the environment, called the `raise_line_contingency_alarm` action. This action is binary and has dimension $(n\_attackable\_lines, 1)$.

As we do not want the operator to be overwhelmed by too many (potentially unnecessary) warnings and raising too many alarms could also harm the trust in the assistant feature, we only want to raise a limited number of alarms. Thus, the number of raisable alarms is limited by a given `attention_budget` α defined as an element of *observation*. Whenever an alarm is raised to require the operator attention, it has a corresponding cost κ. n the other side, if the agent does not require the operator attention, then the "attention budget" increases by $\mu > 0$. Also, to model the fact that human attention is limited, the attention budget is capped by a maximum value A (for example $A = 5$) which ensured that the agent cannot raise more than $\frac{A}{\kappa}$ consecutive alarms. Indeed, it can only raise an alarm if the attention budget is above cost κ. Otherwise it has to wait to recover the necessary budget. In the observation space, along with the `attention_budget` are defined other metrics such as :

- `last_alarm` : the previous alarm
- `is_alarm_illegal` : a boolean information about whether or not the alarm can be raised (for example, an alarm defined in the action space won't be taken into account if the attention budget is too low)
- `time_since_last_alarm` : the number of step since last alarm
- `confidence_rate` : the number of cumulated non-raised alarms averaged over the number passed steps until the current one.

To sump up, at each time step, the alarm is only raisable on each attackable line, in the limit of the given *attention budget* α_t at that time t. The update rule for the attention budget is :

1. $\alpha_{t+1} = \alpha_t \quad \kappa \quad number\_of\_raised\_alarmed_t$ if at least one alarm is raised
2. $\alpha_{t+1} = \alpha_t + \mu$ otherwise

A.7 Customizable environment parameters

The L2RPN challenge offers multiple parameters that could be customized at the creation of the environment during training. At test time however, parameters are fixed with the parameter values listed in the documentation.

An easy parametrization for the environment could be using the following parameters:

```
Difficulty = "0"

NO_OVERFLOW_DISCONNECTION:  true
NB_TIMESTEP_OVERFLOW_ALLOWED: 9999
NB_TIMESTEP_COOLDOWN_SUB: 0
NB_TIMESTEP_COOLDOWN_LINE: 0
HARD_OVERFLOW_THRESHOLD: 9999
NB_TIMESTEP_RECONNECTION: 0
IGNORE_MIN_UP_DOWN_TIME: true
ALLOW_DISPATCH_GEN_SWITCH_OFF: true
ENV_DC: false
FORECAST_DC: false
MAX_SUB_CHANGED: 2
MAX_LINE_STATUS_CHANGED: 2


Difficulty = "challenge" (default)

NO_OVERFLOW_DISCONNECTION:  False
NB_time step_OVERFLOW_ALLOWED: 3
NB_time step_COOLDOWN_SUB: 3
NB_time step_COOLDOWN_LINE: 3
HARD_OVERFLOW_THRESHOLD: 2
NB_time step_RECONNECTION: 12
IGNORE_MIN_UP_DOWN_TIME: true
ALLOW_DISPATCH_GEN_SWITCH_OFF: True
ENV_DC: False
FORECAST_DC: False
MAX_SUB_CHANGED: 2
MAX_LINE_STATUS_CHANGED: 2
```

B Power grid operations

B.1 Physical Variables

Electricity is a form of energy involving the excitement of electrons in metallic elements. In order to develop an understanding of electricity, it is necessary to introduce the fundamental dimension of physical measurement electric charge. Charge is a property of matter arising from atomic structure which is made up of protons (positively charged), electrons (negatively charged) and neutrons (neutral). It is measured in coulombs (C), a charge equal to that of 6.25 10^{18} protons.

Charge induces a force with opposite charges attracting and the same charges repelling. This force creates the ability to produce work and the electric potential or voltage, which is the potential energy possessed by a charge at a location relative to a reference location. It is defined between two points and measured in Volts, denoted with the symbol V. An electric current is a flow of charge through a material, measured in Coulombs per second or Amperes (A) and denoted with the symbol I.

The electrical power is given as the product of the voltage and the current.

$$P = VI \tag{10}$$

Power is measured in Watts, denoted by the symbol W. In order to try to simplify these electrical concepts, an analogy with a physical water system is often used, while not quiet directly analogous, the current is similar to the flow of water in a pipe, say in litres per second. Voltage would be analogous to a height difference, say between a water reservoir and the downhill end of the pipe, or a pressure difference. Intuitively, voltage is a measure of 'how badly the material wants to get there' and current is a measure of 'how much material is actually going'. Power would be analogously produced by the force of water spinning a hypothetical turbine that may rotate a wheel. Intuitively these phenomena are related, increasing the voltage or current in a system increases the power produced. Electrically, this relationship is captured by Ohm's law:

$$V = IR \tag{11}$$

A new variable is introduced here - R - which is the resistance of the material the current is flowing through, analogous to the size of the water-pipe. A smaller pipe makes it harder for large flows and it is the same with current highly conductive materials allowing current to flow easily and poorly conductive materials (called insulators) preventing current from flowing. Whenever an electric current exists in a material with resistance it will create heat. The amount of heating is related to the power P and combining equations 10 and 11 gives:

$$P = I^2R \tag{12}$$

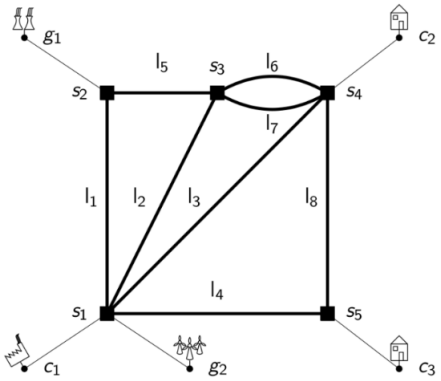

Figure 19: A simple electricity network, showing the circuit nature of a power network, the currents I flowing in the lines and the *interconnectedness* between generators denoted g, customer loads denoted c and substation nodes denoted s.

In order to produce a sustained flow of current, the voltage must be maintained on the conductor. This is achieved by providing a pathway to recycle charge to its origin and a mechanism, called an electromotive force (emf), that compels the charge to return to its original potential. Such a setup constitutes an electric circuit. Again, to oversimplify by relating back to the water analogy - if there is an open pipe in the circuit water will run out. Likewise, if there is a break in an electric circuit, current will not flow but voltage will still be present on the conductor. Simple electric circuits are often described in terms of their constituent components; voltage sources, conductors and resistances. Complex power networks can be described in terms of generation sources, network lines and loads. A simple electrical power network analogous to a simple electric circuit is shown in Figure 19.

Circuit analysis is the goal of estimating the parameters in a circuit given a combination of the voltages, currents and resistances and the fundamental Equations 10, 11 and 12. The more complex the circuit or network, the more complex the analysis will be. Within a circuit, a series of laws known as Kirchhoff's law also help us in the analysis:

- Kirchhoff's voltage law: voltage around any closed loop sums to zero
- Kirchhoff's current law: current entering and exiting any node sums to zero

These principles can be applied at the micro-level to simple circuits, such as plugging in an electric kettle where the element is the resistor or load, the mains outlet is the voltage source and current is proportional to the voltage and resistance of the circuit. Voltage is maintained throughout the circuit when

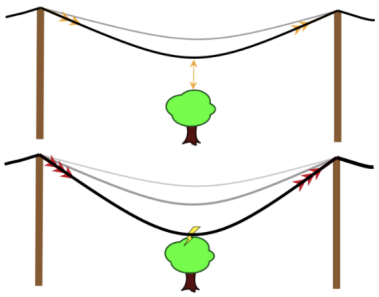

Figure 20: An example of the dangers of overheating power-lines, by transporting too much current, the metallic conductor heats and sags close to the ground causing a flash over to ground and endangering human life.

it is plugged in and current flows from the plug outlet through the wire, into the heating element and back to the plug outlet, completing the circuit. These concepts can also be applied at the macro level, where a house or town could be considered the load and a nuclear power station could be considered the voltage and current source, which is interconnected to the load by power lines. The electricity network is one large circuit, which is constantly satisfying these laws.

B.2 Line thermal limits and congestions

Heating can be desirable. Heating a resistive element is how an electric kettle or heater works. It can also be undesirable - as is the case of power lines - where the heat is energy lost and causes thermal expansion of the conductor making them sag, or fall close to the ground or to people or buildings, as illustrated in Figure 20. In extreme cases, such as a fault condition, thermal heating can melt the wires. As we see from Equation 12 the amount of heating is proportional to the square of the current, so increasing the current has a large effect on the resistive losses. It is for this reason that when electricity is transported over long distances, it is done at high voltages. Based on Equation 11, assuming that the resistance of the line remains constant, to transport the same amount of power, resistive losses are reduced by increasing the voltage and lowering the current. This is the fundamental concept of electricity transmission.

Power will flow from source to load, around the network based on the resistance of the lines in the network. A transmission line has an upper limit to the amount of power that can flow through it before it will fail in service. This limit is given by the thermal properties of the metallic materials, usually copper or aluminium and also cooling and heating due to weather conditions (such as wind, irradiance and ambient temperature). If too much power is forced through the equipment for a long period, and the thermal limits are breached, the equipment is likely to fail in service and be disconnected. In reality, this means overhead lines sag closer to the ground, and may cause flashover as shown

in Figure 20 (the cause of the 2003 blackout in North America) or very expensive equipment such as transformers or cables will be damaged and explode. It is better to disconnect the line than let it sag close to the ground. When the line is disconnected, the same amount of power is still present on the network, but one link has been removed. This means that the power will reroute itself to the new most desirable path based on the resistance, but this rerouting may result in another line or lines being overloaded. The challenge of network operation (and the basis of this RL challenge) is to route the power around the network in the most efficient manner, while avoiding overloads and cascading effects.

B.3 Possible unexpected events on the grid

Contingency can happens in the network, usually the loss of any element on the network (a generator, load, transmission element). Loss of elements can be anticipated (scheduled outages of equipment) or unanticipated (faults for lightning, wind, spontaneous equipment failure). Cascading failures must be avoided to prevent blackouts.

B.4 Operational considerations

The transmission network is controlled from a control centre, with remote observability from this centre to all transmission network elements. The network operators can control most network elements such as lines and substations via remote control command. The control centre also has visibility of all the generation sources and all the loads. Generation is controlled by the control centre operator sending dispatch instructions to change the outputs. Some loads can be controlled by the control centre, but, in general, the distribution system operators control switching of the load. For small to medium-sized countries, usually, there is one control centre with responsibility for network control but for larger countries like the USA, Canada there are multiple control centres that control the network at a state or regional level on a functional basis. These control centres coordinate their activities with their neighbouring control centres.

The network operator's role is to monitor the electricity network 24 hours per day, 365 days per year. The operator must keep the network within its thermal limits, its frequency ranges and voltage ranges for normal operation and contingency state operation as described above. For normal operation, the operator has a range of actions at their disposal to manage the network within its constraints, such as switching, generator dispatch and load disconnection. For the contingency state operation, the operator must act ahead of time to mitigate contingencies that may occur for the unexpected loss of any single element, using the prescribed range of actions. The operator must also plan the network operation for the loss of any element for a scheduled outage for maintenance. The network must operate securely for the entirety of the planned outage, not just for the moment the outage is taken. The operator must plan for and manage the network within its limits at the system peak, i.e. the largest

load demand of the day, the generation must be managed so that the generation load balance (measured by the frequency) is maintained at the peak of the day.

The power network is operated by ensuring the three primary constraints are met at all times, in all areas of the network.

- Thermal limits of transmission equipment are not breached (measured in current with units of Amperes (A) or power with units MegaWatts (MW)).
- Voltage maintained within a defined range (measured in voltage, units of Volts (V)).
- Generation and load balanced at all times (measured in power, units of Megawatts (MW). The balance between load and generation is approximated by frequency measured in Hertz (Hz).

Operators also have to consider other operational rules such as a 1 to 3 number of actions from 5 to 20 minutes to limit the risk of human errors or action failure. There are also cooldown times of 15 minutes or more to switch again some breakers, as well as maximum 15 minute time-delay before overload line-tripping by protections.

B.5 Cost of operations details

Energy Losses Cost We will recall that transporting electricity always generates some energy losses[15] $\boldsymbol{\mathcal{E}_{loss}(t)}$ due to the Joule effect in resistive power lines at any time t:

$$E_{loss}(t) = \sum_{l=1}^{n_l} r_l \quad y_l(t)^2 \tag{13}$$

At any time t, the operator of the grid is responsible for compensating those energy losses by purchasing on the energy market the corresponding amount of production at the marginal price $\boldsymbol{p(t)}$. We can therefore define the following energy loss cost $\boldsymbol{c_{loss}(t)}$:

$$c_{loss}(t) = E_{loss}(t) \quad p(t) \tag{14}$$

Topological action can increase or decrease $E_{Loss}(t)$. This already leads to a continuous optimization problem to solve.

Flexibility Cost Then we should consider that operator decisions when taking an action can induce costs, especially when requiring market actors to perform specific actions, as they should be paid in return. Topological actions are mostly free, as the grid belongs to the power grid operator, and no energy cost is involved. However, redispatching actions involve producers which should get

[15] This energy loss corresponds to 2.2% of the total energy consumption on high voltage power grids: https://bilan-electrique-2018.rte-france.com/loss-rate/?lang=en

paid. As the grid operators ask to redispatch energy $\mathcal{E}_{redispatch}(t)$ or curtail energy $\mathcal{E}_{curtailment}(t)$ from solar or wind farms, some power plants will increase their production by $E_{redispatch}(t)$ while others will compensate by decreasing their production by the same amount to keep the power grid balanced. Hence, the grid operator will pay both producers for this redispatched energy at a cost $c_{redispatching(t)}$ higher than the marginal price $p(t)$ by some factor α. Also agents can use batteries with energy $E_s(t)$ produced or consumed under a fixed 10€/MWh, not driven by market prices:

$$\begin{aligned} c_{flexibility}(t) &= c_{redispatching}(t) + c_{curtailment}(t) + c_{storage}(t) \\ &= 2\alpha p(t)(E_{redispatch}(t) + E_{curtailmenet})(t) + 10 \sum jEs(t)j,\ \alpha \geqslant 1 \end{aligned} \tag{15}$$

Blackout cost In case of a blackout, the cost $c_{\text{blackout}}(t)$ at a given time t would be proportional to the amount of consumption not supplied Load(t), at a price higher than the marginal price $p(t)$ by some factor β:

$$c_{\text{blackout}}(t) = \text{Load}(t) \quad \beta \quad p(t),\ \beta \geqslant 1 \tag{16}$$

Notice that Load(t) $>>$ $E_{\text{redispatch}}(t), E_{\text{loss}}(t)$ which means that the cost of a blackout is a lot higher than the cost of operating the grid as expected. It is even higher if we further consider the secondary effects on the economy[16]. Furthermore, a blackout does not last forever and power grids restart at some point. But for the sake of simplicity while preserving most of the realism, all these additional complexities are not considered here.

B.6 Upcoming Operational Challenges for an assistant

In the wake of the current energy transition, along with economy and technology shifts, the power grid is subject to new changes. As described in Figure 21, many changes outside the grid are having significant operational impacts. For instance, digitization of procedures, social networks, new media consumption patterns, adaptation to climate change, are modifying how end-users consume electricity. These new elements create new electricity usages, eventually leading to new grid dynamics. Moreover, the grid itself is evolving with more micro grids, renewable energies, distributed power plants and storage capacities connected to the grid. From a market perspective, higher energy prices, new market participation and mechanisms, the sharing of open data and information are having a significant impact on power grid operations.

Thus, the grid is increasingly pushed to its operational limits where more congestions, along with higher operational uncertainty and faster evolutions of the system dynamics occur. To keep operating the grid safely in the future,

[16] More information can be found on this blackout cost simulator: https://www.blackout-simulator.com/

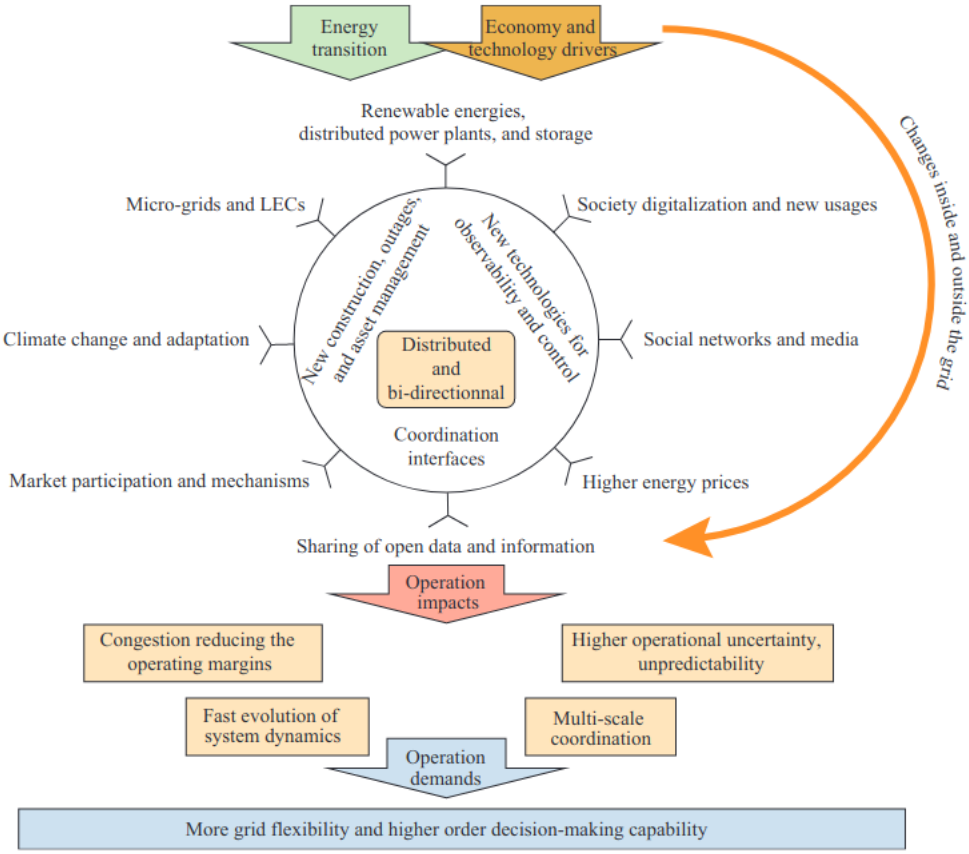

Figure 21: New operational needs under energy transition that is impacting operations along different dimensions. (Image from [12])

we need to have more flexibility in operations and a better and faster decision-making process.

In this new context, the current grid monitoring trough various screens in control room (see Figure 2) is not sustainable. We need to shift from the supervision paradigm, where a lot of information are displayed to the hypervision one where synthetic information are displayed to the "right person at the right time" of their decision process. To do so, in a near future, all theses new grid operational information and actions could be centralized using an assistant as a unified interface. This central element of the future grid operation could either gather information from various sources, make recommendations to the operator, and contextualize a given situation. The role of future assistants has been further detailed in section 4.

C Available Operational Flexibilities

The powergrid operator's role is to monitor the electricity network 24 hours per day, 365 days per year. The operator must keep the network within its thermal limits, its frequency ranges and voltage ranges for normal operation and contingency state operation as described above. For *normal operation*, the operator has a range of actions at their disposal to manage the network within its constraints, such as switching, generator dispatch and load disconnection. For the *contingency state operation*, the operator must act ahead of time to mitigate contingencies that may occur for the unexpected loss of any single element, using the prescribed range of actions. The operator must also plan the network operation for the loss of any element for a scheduled outage for maintenance.

Constraints on the system, such as line congestions are alleviated in real-time by powergrid operators using a range of remedial actions, from least to most costly as follows:

- **Switching lines** on the network in or out
- **Splitting or coupling busbars** at substations (see Figure 22) together. This means a node can be split into two elements or connected together as a single element
- **Redispatch generation** to increase or reduce flows on lines
- **Load shedding** disconnecting some load from the grid

From a cost perspective, the disconnection of load should be avoided due to the disruption to society, business and daily life. Redispatching generation can also be expensive. The electricity is managed by a market, based on the cost per unit of energy supplied. If the network operators need to redispatch expensive generation, this can be sub-optimal from a market perspective and cause increased costs to customers. To provide operational flexibility, substations are usually designed so that they can be separated into two or more constituent

Figure 22: 3 categories of flexibility with different cost and spread

parts. Coupling a substation can serve to reroute power in a network and is an option to alleviate line overloads. Switching lines and coupling busbars at substations are the least costly option to alleviate thermal overloads on the network. There is considerable operational flexibility that is under-utilized on power networks that can be released by switching actions and topology changes. This network flexibility is easy to implement and the least costly option. One of the goals of the RL challenge is to explore the range of switching options available and to utilize topology changes to control power on the network.

The network operators can control most network elements such as lines and substations via remote control command. The control centre also has visibility of all the generation sources and all the loads. Generation is controlled by the control centre operator sending dispatch instructions to change the outputs. Some loads can be controlled by the control centre, but, in general, the distribution system operators control switching of the load. For small to medium-sized countries, usually, there is one control centre with responsibility for network control but for larger countries like the USA, Canada there are multiple control centres that control the network at a state or regional level on a functional basis. These control centres coordinate their activities with their neighbouring control centres. In France, there were 8 of them , now reduced to 3 but with other kind of control room appearing.

References

[1] Artelys, Armines, and Energies Demain. A 100% renewable electricity mix? analyses and optimisations. 2015.

[2] Richard Bellman. A markovian decision process. *Indiana Univ. Math. J.*, 6:679–684, 1957.

[3] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym, 2016.

[4] B. Donnot. Grid2op- A testbed platform to model sequential decision making in power systems. . https://GitHub.com/rte-france/grid2op, 2020.

[5] Matthias Dorfer, Anton R Fuxjäger, Kristian Kozak, Patrick M Blies, and Marcel Wasserer. Power grid congestion management via topology optimization with alphazero. *arXiv preprint arXiv:2211.05612*, 2022.

[6] Anton R Fuxjäger, Kristian Kozak, Matthias Dorfer, Patrick M Blies, and Marcel Wasserer. Reinforcement learning based power grid day-ahead planning and ai-assisted control. *arXiv preprint arXiv:2302.07654*, 2023.

[7] 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.

[8] A Marot, B Donnot, S Tazi, and P Panciatici. Expert system for topological remedial action discovery in smart grids. 2018.

[9] Antoine Marot, Benjamin Donnot, Karim Chaouache, Adrian Kelly, Qiuhua Huang, Ramij-Raja Hossain, and Jochen L Cremer. Learning to run a power network with trust. *Electric Power Systems Research*, 212:108487, 2022.

[10] Antoine Marot, Benjamin Donnot, Gabriel Dulac-Arnold, Adrian Kelly, Aidan O'Sullivan, Jan Viebahn, Mariette Awad, Isabelle Guyon, Patrick Panciatici, and Camilo Romero. Learning to run a power network challenge: a retrospective analysis. In *NeurIPS 2020 Competition and Demonstration Track*, pages 112–132. PMLR, 2021.

[11] 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. In *PSCC2020 (preprint)*, 2020.

[12] Antoine Marot, Adrian Kelly, Matija Naglic, Vincent Barbesant, Jochen Cremer, Alexandru Stefanov, and Jan Viebahn. Perspectives on future power system control centers for energy transition. *Journal of Modern Power Systems and Clean Energy*, 10(2):328–344, 2022.

[13] Antoine Marot, Alexandre Rozier, Matthieu Dussartre, Laure Crochepierre, and Benjamin Donnot. Towards an ai assistant for power grid operators. In *HHAI2022: Augmenting Human Intellect*, pages 79–95. IOS Press, 2022.

[14] Loïc Omnes, Antoine Marot, and Benjamin Donnot. Adversarial training for a continuous robustness control problem in power systems. In *2021 IEEE Madrid PowerTech*, pages 1–6, 2021.

[15] Adrien Pavao, Isabelle Guyon, Anne-Catherine Letournel, Xavier Baró, Hugo Escalante, Sergio Escalera, Tyler Thomas, and Zhen Xu. Codalab competitions: An open source platform to organize scientific challenges. *Technical report*, 2022.

[16] 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.

[17] 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.

[18] Richard S Sutton and Andrew G Barto. *Reinforcement Learning: An Introduction.* A Bradford Book, second edition, 2018.

[19] Jan Viebahn, Matija Naglic, Antoine Marot, Benjamin Donnot, and Simon H Tindemans. Potential and challenges of ai-powered decision support for short-term system operations. In *CIGRE Session 2022*, 2022.

[20] Bo Zhou, Hongsheng Zeng, Yuecheng Liu, Kejiao Li, Fan Wang, and Hao Tian. Action set based policy optimization for safe power grid management. In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13{17, 2021, Proceedings, Part V 21*, pages 168–181. Springer, 2021.