IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 5 March 2024, accepted 13 March 2024, date of publication 1 April 2024, date of current version 19 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3383442

APPLIED RESEARCH

Improving the Computational Efficiency of the Unit Commitment Problem in Hydrothermal Systems by Using Multi-Agent Deep Reinforcement Learning

PHILIP GUERRA, (Student Member, IEEE), ESTEBAN GIL, (Member, IEEE), AND VÍCTOR H. HINOJOSA, (Member, IEEE)
Department of Electrical Engineering, Universidad Técnica Federico Santa María, Valparaíso 2390123, Chile
Corresponding author: Esteban Gil (esteban.gil@usm.cl)

This work was supported in part by Chilean National Agency for Research and Development (ANID) under Grant Basal FB0008 and Grant Fondecyt 1231892; and in part by Universidad Técnica Federico Santa María, Chile, under Project USM PI_LIR_2022_03.

ABSTRACT In power systems with a significant hydroelectric component, instances of the Unit Commitment (UC) problem may be much more computationally intensive due to the longer decision horizons and the additional hydro constraints. Therefore, this paper presents a methodology to reduce the solution space to accelerate 168-hour-ahead UC formulated as a Mixed-Integer Linear Program (MILP). First, an offline model maps environment observations to actions in a Multi-Agent Deep Reinforcement Learning (MADRL) model. This mapping uses historical power system operation data to determine the on/off status of specific generation units. Then, the online model uses the binary variable solutions obtained by the offline model to solve a UC problem with a reduced solution space. The Multi-Agent approach allows each agent, based on Artificial Neural Networks (ANN) with a Temporal Convolutional Network (TCN) architecture, to group units that are located in the same region. A shared cumulative reward function is used to adjust simultaneously the different ANN weights during the learning phase. The effectiveness of our method is demonstrated using real operational data of the Chilean National Electricity System, achieving statistically significant lower computation times and a negligible error that is within the integrality gap of the solver.

INDEX TERMS Artificial neural networks, multi-agent deep reinforcement learning, unit commitment, variable reduction.

I. INTRODUCTION

Short-term generation scheduling involves minimizing the operation costs in a specific horizon by defining an optimal schedule for the different generators while satisfying numerous physical, operational, and economic constraints. This is a much more difficult task for systems with an important hydroelectric component due to the limits of water reservoirs and the stochastic nature of hydro inflows. Specifically, in hydrothermal systems such as the Chilean National Electricity System, the Unit Commitment (UC) problem entails optimizing the sequence of on/off status for thermal generators, requiring solutions for horizons extending up to 168 hours. The complexity deepens with the growing penetration of Variable Renewable Energy (VRE) sources and the inclusion of new technologies and novel constraints. As renewable generators are included as variables to model transmission limits due to curtailment effects, more decision variables are added to UC formulations, adding even more time to solve the associated Mixed-Integer Linear Programs (MILP). Thus, power systems operational planning tools in the energy transition require balancing the growing computational demands with solution quality.

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

There have been many proposals to improve the computational performance in the UC problem focused on using historical power system operation records. A review of current machine learning (ML) trends applied to the UC problem is shown in [1]. The first works incorporating the use of ANN to solve the UC problem are from the 1990s. In [2] and [3], ANNs are used to solve the UC problem in a simplified way. In [2], a neural network uses the load profiles as input and thermal generators' on/off status as outputs (26 units). In [3], Hopfield networks are used for a system with 17 thermal units. Reference [4] introduces a model based on a three-layer fully connected network model to estimate the units' status by utilizing nodal load profiles as input. Subsequently, the UC problem is solved using a simulated annealing method, a heuristic optimization algorithm to find near-optimal solutions to combinatorial optimization problems. In these works, good results are achieved related to improving resolution times, but they are only tested in small test systems with few thermal units.

More recent works have tried to use data-driven approaches to find patterns in the optimization problem and treat it as a classification/clustering problem. In [5], a reduction of variables to the classic MILP formulation is proposed for the Security-constrained UC (SCUC) problem. With an offline/online scheme, the offline model obtains the solutions of the binary variables, whereas the online model solves the reduced SCUC problem with the MILP formulation. This paper uses the k-means method to classify the net load demand in the buses and find feasible solutions for the state of the units.

Another trend seen in recent years is the application of ML techniques to improve the performance of MILP solvers based on Branch & Bound. In [6], ML is used to improve the branching strategy. The approach involves emulating the decisions made by an effective branching strategy, specifically strong branching, through a fast approximation. This approximation is generated using an ML technique based on a collection of observed branching decisions derived from strong branching. The proposed approach involves extracting features to represent the state of a potential branching variable within a specific tree node. Comprehensive strong branching decisions are obtained by solving a set of training instances, and subsequently, a regressor is trained to predict the estimated branching score.

In [7], three strategies based on ML are proposed to extract features from the SCUC problem and solve the reduced problem using the MILP formulation. On the one hand, they utilize the k-nearest neighbors (kNN) algorithm to determine which constraints should be initially included in the first iteration of the UC resolution and which ones should be excluded. On the other hand, the authors propose a solution predictor using the kNN method as a warm start for the MILP solver to obtain initial feasible solutions. Finally, they introduce a Support Vector Machine (SVM)-based affine subspace predictor to reduce the search area. The predicted results potentially enhance the running time and solution quality of the MILP solver. Solution times are reduced 4.3 times on systems of 1888 and 6515 buses.

Some studies have examined different ML algorithms to predict units' on/off status, including feasibility techniques to assess solution viability. These studies have also utilized demand profiles as inputs to train the ML algorithms. For example, in [8], logistic regression (LR), deep neural networks (DNN), random forest (RF), and kNN algorithms are utilized to obtain a reduced SCUC problem. The commitment schedules' accuracy is validated using various test cases, such as the IEEE 24-bus, IEEE 73-bus, IEEE 118-bus and South Carolina Synthetic Grid 500-bus. Authors also incorporate a feasibility layer to address infeasible solutions, demonstrating high training accuracy. The results displayed significant speed-up ratios ranging from 3.4 to 3.6 on average for all test systems. In addition, tests for the Polish 2383-bus system show a speed-up ratio of 6.9.

Similarly, in [9], experiments are conducted on the IEEE 24-bus system and a practical 5655-bus system using kNN, DNN, RF and Decision Trees (DT) algorithms. On average, each algorithm achieved a computational improvement of at least 40% while maintaining the cost variation within 1%. Compared to the other algorithms, kNN consistently achieved notable computational enhancements without sacrificing optimality in any scenario. Based on the results obtained by various models, the authors concluded that learning the on/off status of the units may not necessitate highly sophisticated models. In [10], the authors focused on analyzing historical power system data patterns, incorporating a model based on Graph Neural Networks (GNN) and long short-term memory (LSTM) layers. They validated their findings on various power system test cases, including the IEEE 24-bus, IEEE 73-bus, IEEE 118-bus, and South Carolina Synthetic Grid 500-bus system. One of the significant contributions of their research is the application of the GNN layer, harnessing the advantages of graph-based structures. The proposed approach displayed significant time savings, with reductions ranging from 20% to 50%, depending on the specific study case.

The aforementioned works demonstrate a common approach in training ML models using load profiles or net demand as inputs. These inputs consider consumers' electricity consumption patterns and consider the contribution of renewable energy sources to electricity generation. However, those works have focused on thermal systems without considering systems with a strong hydroelectric component, like the Chilean electric system. Moreover, research has demonstrated that integrating domain knowledge into machine learning models can significantly enhance model performance. For example, [11] propose two clustering algorithms based on domain knowledge, illustrating their advantages compared to 13 other clustering algorithms. In hydrothermal systems, operational scheduling is intricately linked to water reservoir management, introducing unique challenges to the UC problem, such as longer horizons

and an expanded solution space. The operation of thermal power plants within these systems relies on managing water reservoirs. Incorporating domain-specific information, such as reservoir generation, into the model can augment the model's capacity to accurately determine the optimal operation of thermal units in response to dynamic system conditions.

Furthermore, this is closely linked to the fact that most works consider ML models, such as kNN, RF or SVM [7], [8], [10], to treat the problem as one of classification. However, despite its excellent performance in other sequential decision-making problems, the use of Reinforcement Learning (RL) to obtain relevant decision variables in the optimization problem has not been explored. By employing a Multi-Agent Deep Reinforcement Learning (MADRL) framework to tackle the UC problem, we introduce a novel theoretical and methodological perspective to power system optimization, demonstrating how advanced machine learning techniques can significantly enhance the efficiency and reliability of hydrothermal power system operations.

Recent advances in ML techniques have shown the potential of RL techniques combined with ANN for sequential decision-making. This integration effectively resolves real-world complex problems by facilitating interactions with the environment and learning from reward signals. For example, in [12], an offline/online scheme is proposed to obtain power levels dispatched together with voltage levels in an Optimal Power Flow (OPF) problem. RL is used to offer a cost function that includes operational constraints. In [13], a MADRL scheme is proposed, based on ANN and RL, to operate a hybrid photovoltaic plant with energy storage participating in electricity and ancillary service markets. Two ANN-based agents are proposed for the day-ahead and real-time markets. Both networks are trained under the same reward function so that the weights of both networks are adjusted simultaneously. A similar approach is adopted in this study, as each agent can specialize in solving a specific problem within the context of the overall task, and agents can collaborate and share knowledge during the training process. This application of MADRL to hydrothermal UC problems not only showcases the adaptability of machine learning models to complex energy systems but also enriches the theoretical foundations of reinforcement learning, particularly in multi-agent settings where cooperative learning strategies are critical. Hence, this approach enhances the global learning of the system and can lead to more efficient and robust solutions.

From a methodological point of view, the MILP-based UC formulation is solved by actual commercially available solvers obtaining high-quality UC solutions in adequate simulation time. However, this study significantly enhances the resolution process of the UC problem in a real hydrothermal system. The MADRL algorithm allows us to take advantage of its capabilities for sequential decision-making in complex environments. In the proposed MADRL model, several agents are trained short-term generation scheduling results conducted by the Independent System Operator (ISO) of the Chilean National Electricity System. After setting the hyper-parameters with a validation set, out-of-sample results are contrasted against actual UC results obtained by the ISO to determine the performance and convergence conditions of the introduced methodology solving real day-ahead conditions. Therefore, the main contribution of this work is to propose a MADRL-based offline scheme to determine specific thermal generation units' on/off status reducing the solution space, and to solve instances of MILP-based UC problems applied to a real-life large-scale hydrothermal power system. Simulation results demonstrate that the proposed method significantly accelerates the resolution time remaining within the established error margins so that the mathematical framework is under study for its potential implementation by the Chilean ISO. Bearing in mind (i) the interest of security-constrained unit commitment models, (ii) the relevance of modeling hydrothermal power systems, (iii) the introduced MADRL approach, (iv) the modeling improvements described in the paper upon the state-of-the-art security-constrained unit commitment, and (v) the successful numerical experience applied to a real-life hydrothermal power system reported in the manuscript, we feel that the contents of this paper constitute an original and substantial contribution to the technical knowledge. Here, we underscore our theoretical contributions: to the field of machine learning, by advancing MADRL methodologies for complex decision-making environments; and to power system optimization, by presenting novel solutions that significantly reduce computational burdens while addressing the unique challenges of hydrothermal systems.

The rest of this paper is organized as follows: Section II introduces the required background in UC and hydrothermal generation scheduling. Section III proposes the MADRL framework to solve the problem. The case study is presented in Section IV. Section V presents numerical results regarding solution quality and computational performance. Finally, Section V concludes this paper.

II. BACKGROUND

A. MILP FORMULATION OF UC

The UC problem is a family of optimization problems dealing with scheduling power-generating units in a specific period. The UC solution must respect many operational constraints. Equation 1 presents a generic formulation for the UC problem:

$$\min_{\boldsymbol{x},\boldsymbol{y}} \ \boldsymbol{c}^T\boldsymbol{x} + \boldsymbol{d}^T\boldsymbol{y} \tag{1a}$$

$$\text{subject to } A\boldsymbol{x} \leq \boldsymbol{b}, \tag{1b}$$

$$H\boldsymbol{y} \leq \boldsymbol{h} \tag{1c}$$

$$G\boldsymbol{x} + E\boldsymbol{y} \leq \boldsymbol{g} \tag{1d}$$

$$\boldsymbol{x} \in \{0,1\}^{|\mathcal{G}|\times|\mathcal{T}|} \tag{1e}$$

The vector $\boldsymbol{x}$ represents the binary variables related to the on/off status of the thermal power plants for the planning horizon period. The vector $\boldsymbol{y}$ represents the continuous variables related to each unit's dispatched power and reserves. The sets $\mathcal{G}$ and $\mathcal{T}$ contain the indices that identify each generator and the scheduling-time period, respectively. The cost function and constraints are written linearly, resulting in a MILP formulation.

The objective function seeks to minimize the production costs, considering the fuel costs and start-up and shutdown costs, in addition to the costs of energy and reserves not served. The term $\boldsymbol{c}^T\boldsymbol{x}$ represents the generators' start and shutdown costs. The term $\boldsymbol{d}^T\boldsymbol{y}$ corresponds to the dispatch costs and the unsupplied energy costs. The constraint (1b) includes the minimum-up and minimum-down to prevent starting and shutting down too frequently. The constraint (1c) contains the ramping rates of the generators, and the constraint (1d) includes the power balance requirements, unit reserve and their generation limits. It should be noted that the problem can also consider restrictions associated with VRE units, transmission and security constraints using a linear DC approximation [14].

B. MILP-BASED UC CHALLENGES IN HYDROTHERMAL SYSTEMS

The UC problem in hydrothermal systems is much more complex than for purely thermal systems because of the need to manage hydro energy storage of water reservoirs, hydro cascading constraints, and considerations for alternative water uses. For example, the UC problem includes many complex irrigation constraints in the Chilean system.

Also, the operational planning horizons are much longer because of the limited storage capacity of hydro reservoirs. Whereas in thermal UC, the horizon extends to 1 day or, at most, a few days ahead, in many hydrothermal systems, the horizon extends to a week, increasing the number of variables involved [15]. In this paper, the MILP formulation uses an operational planning horizon of 168 hours, where the binary variables constraints of the four last days are relaxed.

Furthermore, UC in hydrothermal systems with large reservoirs needs to use results from hydro coordination models (usually based on Stochastic Dual Dynamic Programming models, SDDP [16]), which estimate the opportunity costs of water over much longer horizons.

III. PROPOSED MADRL-BASED SOLUTION-SPACE REDUCTION FRAMEWORK FOR UC PROBLEM

In this study, while acknowledging the existence of various methodologies aimed at optimizing the UC problem, our primary focus is on the practical application and implementation of a MADRL approach specifically tailored to the Chilean hydrothermal power system, offering a direct, real-world comparison with the current UC strategies employed by the Chilean ISO, rather than a theoretical exploration of methodological alternatives.

A. OFFLINE/ONLINE PROPOSED FRAMEWORK

In this paper, an offline model maps environment observations to actions using a MADRL model, determining specific generation units' on/off status that are then fed to an online MILP-based model. Fig. 1 shows a simplified scheme of the offline/online proposed framework.

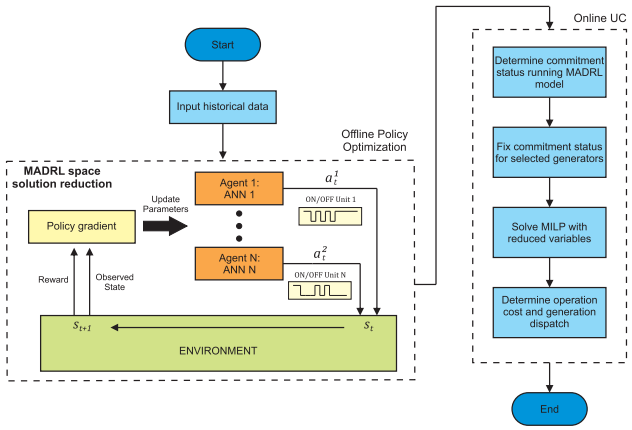

FIGURE 1. Flowchart of the proposed framework.

In the offline stage, the ANN-based agents of the MADRL model are trained through interaction with the simulated environment, built with historical system operation scheduling data. While reinforcement learning can be computationally expensive to train, once the MADRL model is fine-tuned, its predictions about the commitment status of the selected generators are nearly instantaneous. In the online stage, these predictions can then be used to reduce the solution space of the security-constrained UC, which is where computational efficiency improvement is sought.

B. MADRL

MADRL is chosen for this operational power system problem due to its suitability for sequential decision-making and its capabilities for adapting to changing environments [17], [18], [19]. In this work we introduce a MADRL model that explores the relationship between historical data and UC solutions. The aim is to significantly decrease the computational burden of the MILP-based UC problem by reducing the solution space.

In RL, the learning agent interacts with an environment and strives to deduce the correct output. Although it receives feedback on the quality of its actions, unlike in supervised learning, the correct output is not explicitly disclosed (and the feedback may be delayed). Learning in this context is driven by exploration, involving trial and error. Thus, in our proposed model, the agents only receive feedback on the quality of their actions regarding how close the proposed solution was to the solution of the MILP-based hydrothermal UC obtained by PLEXOS. However, it is never informed what that solution was, and the feedback is provided after the action. To elaborate on this, the approach involves utilizing historical data to simulate the decision-making

environment for agents regarding the on/off status of units. This simulation occurs during the learning phase, where the environment provides states (s_t) encompassing inputs like demand, VRE, reservoir levels, and unit on/off status. Moreover, the environment quantifies the reward for the agent's actions. However, it does not provide the right answer.

To refine the policies of the ANN-based agents, we leverage historical power system records, encompassing previous binary variables of the UC solutions, hydro generation scheduling for the reservoirs, nodal loads, VRE generation, and a temporal representation. Hydro generation scheduling in the reservoirs holds significance, as it influences the scheduling of thermal units in hydrothermal systems, contingent on water usage. This comprehensive dataset enables us to scrutinize temporal patterns and trends in the operation of the units.

Then, multiple agents are applied to determine the on/off status of the different units. The number of agents depends on the size of the system and its topology. The agents (ANNs) make hourly decisions (a_t) based on observations of the environmental state (s_t) and update the weights of the ANNs using a policy gradient. The decisions a_t represent the on/off status of the different units, and the environmental state s_t are the inputs, i.e., previous on/off status of the units, VRE generation, reservoir scheduling, nodal load and time representation. Finally, the training phase utilizes a shared reward function to encourage agent collaboration.

Subsequently, the online procedure solves the UC using a standard solver while fixing the commitment variables in certain generation units obtained from the MADRL offline policy optimization. This procedure reduces the number of binary variables in the MILP problem, hence decreasing the computational burden. With the UC solution, we can then report operational costs, hydro energy, and dispatch levels for each generation unit.

An alternative approach to using multiple agents would be a single neural network under a multi-task learning scheme [20], saving computation at inference time as only a single network would need to be evaluated. Unfortunately, this often leads to inferior overall performance, as task objectives can compete [13]. Furthermore, [21] empirically demonstrates that the loss or gain of performance depends on the relationship between the jointly trained tasks. Intuitively, the features and behaviour of different generators are different enough to potentially compromise generalization performance. Therefore, we use separate agents to predict the on/off status of different units or groups of units of similar characteristics.

C. FEATURE SELECTION

One of the main advantages of ANNs is their ability to learn non-linear and complex relationships so that we use them to map the prediction policy of our agents based on historical data. As illustrated in Fig. 2, each ANN-based agent is fed with time series data representing the relevant system's operation variables, representing the system's state (s_t). The literature recommends that the best features representing the system's state are those changing in time, as they can capture the dynamics of power system operations [2], [5], [8], [9], [10].

D. INPUT VARIABLE SELECTION

The ISO's experience is also considered for selecting the relevant input variables. For our implementation, we used nodal load profiles, VRE power scheduling, and the different units' on/off status records as inputs of each ANN. Notice that load and VRE can be local or regional locations. Finally, we add a time representation of the hour of the day using integer numbers.

Furthermore, as demonstrated in [11], incorporating domain knowledge into machine learning problems can significantly enhance the performance of algorithms. The UC problem is different in hydrothermal systems, and considering the ISO's experience, the generation schedules of hydro units with significant water reservoirs are also considered input variables. Thus, the operational dynamics of thermal power plants exhibit considerable variation between wet and dry months. While incorporating hydrological conditions into the model's inputs can somewhat mitigate this issue, it is crucial to conduct training and testing of the model across different months. This approach enables an examination of the inherent variations between wet and dry periods. Additionally, the emergence of atypical hydrological conditions might necessitate retraining of the agents to maintain model accuracy and relevance.

E. ANN ARCHITECTURE

An ANN-based agent is illustrated in Fig. 2. We use a sequence of multiple features to map the on/off status. Although in power system time series forecasting, it is common to use LSTM layers (introduced in [22]), in this work, we prefer to employ temporal convolutional networks (TCN) [23]. This novel family of architectures has demonstrated superior performance to LSTM in various tasks, primarily due to their ability to maintain a longer effective memory. TCNs are based on several important characteristics. In the first place, they use causal convolutions, which allow the output at time t to be convolved only with elements from time t and earlier in the previous layer. This allows an output of the same length as the input and no leakage from the future into the past. The second characteristic is the use of dilated convolutions, allowing an exponentially large receptive field so that a top-level output can represent a wider range of inputs. Also, TCN, unlike recurrent networks such as LSTM, uses parallel convolutions to process the input sequences, which is faster than the sequential processing of recurrent networks. Finally, to deal with deeper architectures, it is common to concatenate several TCN blocks, including residual connections [24] to improve the learning procedure

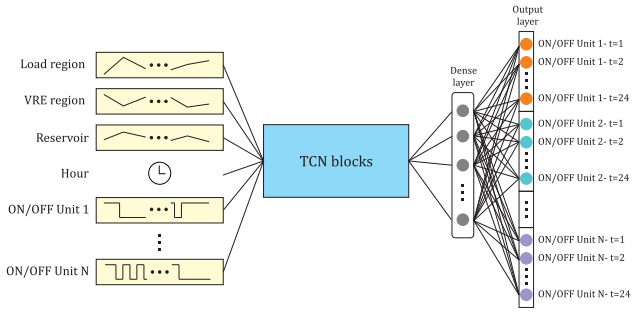

FIGURE 2. Proposed ANN-based agent.

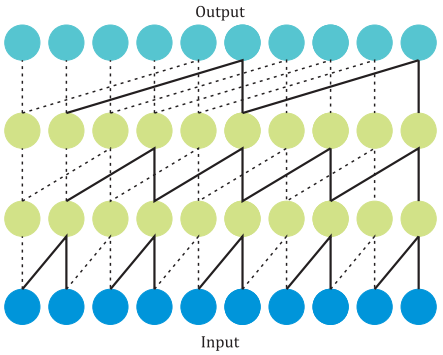

FIGURE 3. TCN block with three convolutional layers with kernel size two and dilations [1,2,4].

in deep architectures with many parameters. Fig. 3 illustrates a TCN block.

The TCN block's output is fed into a dense layer with *ReLU* activation functions. The output layer consists of the on/off status decisions. The output length depends on how many units it tries to predict and the horizon length. For example, if the horizon prediction is $|T|$ hours, and the agent decides the statuses of three generation units, the output layer will consist of $3 \times |T|$ units. The activation function for the output layer is a *sigmoid*(), as this function domain is [0, 1], which represents the status of the units.

F. LEARNING PHASE

As is usual in time series manipulation in ML, we split our data into three consecutive groups: training, validation, and testing. We fit the ANN weights using the training set after every iteration. Next, we use the validation set to select the best model considering the different hyper-parameters of the ANNs. Finally, the test set evaluates the fitter model with out-of-sample data.

One of the keys of the multi-agent model is to achieve a collaborative learning phase between the ANN-based agents using a shared cumulative reward function. In this case, the outputs of the different neurons are rounded to 0 or 1 because of the on/off status of the units. This is similar to a multi-label classification problem (MLCP) [25], which predicts multi mutually non-exclusive classes. In our problem, we decided that every agent will have a binary cross-entropy function as a reward function. This type of function is commonly used to compare the predicted probabilities to actual class output in MLCP. The cross-entropy function is zero when the predicted and real values are equal. When they are not, their value depends on how close or far they are. The reward function for a particular agent is obtained as follows:

$$r_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \cdot \log a_i + (1-y_i)\cdot \log(1-a_i)\right] \quad (2)$$

where y_i is the real value, a_i is the decision made by the agent, and n the number of cases. The average cumulative shared reward function is obtained as follows:

$$R = \frac{1}{N}\sum_{j=1}^{N} r_{BCE,j} \quad (3)$$

This is the average of the reward functions of the different agents in the model. This function depends on the weights of every agent. The gradient is calculated over the UC decisions made by every ANN-based agent. It is important to mention that the agents do not share weights between them but are influenced by each other because of the shared reward function used in the learning phase. We also use mini-batches for the training phase for a computationally more efficient process. We update each ANN's weights using back-propagation in every training iteration. The gradient of each ANN can be decomposed as follows:

$$\nabla_\theta R = \nabla_{\boldsymbol{a}} R \cdot \nabla_\theta \boldsymbol{a} \quad (4)$$

where $\nabla_{\boldsymbol{a}} R$ is the gradient of the cumulative reward function concerning the actions $\boldsymbol{a}$, i.e. the status of the units in the horizon $\boldsymbol{a} = [\boldsymbol{\mu}_{n=1,t}, \ldots, \boldsymbol{\mu}_{n=N,t},]$, and $\nabla_\theta \boldsymbol{a}$ is the gradient of $\boldsymbol{a}$ with respect to neural network parameter θ. Algorithm 1 describes the learning phase of the offline policy optimization block.

Algorithm 1 Learning Phase

```
Randomly initialize critic ANN's weights
Initialize replay buffer R
for each training iteration do
    Randomly sample a mini-batch of replay buffer R
    Receive initial observation state S_0
    for t = 1 to T do
        Determine action a_t
        Collect UC solutions from ANNs
    end for
    Observe state and loss function
    Calculate the gradient ∇_θ R
    Update ANN's weights using policy gradient
    if stop criterion is met then
        break
    end if
end for
```

IV. CASE STUDY

The proposed framework is tested in the Chilean National Electric System. As of December 2022, it has an installed generation capacity of 33,218 MW. Renewable sources account for 62.0% of the installed capacity (22.3% hydroelectric, 24.1% solar, 13.0% wind, 2.3% biomass, and 0.3% geothermal), whereas thermal sources account for 38.0% (13.0% coal, 15.1% natural gas, and 9.8% oil) [26]. The training experiments for the ANNs are conducted on a machine equipped with an Intel(R) Xeon(R) CPU @ 2.20GHz, NVIDIA A100-SXM4-40GB GPU, and 85 GB of RAM available. The MILP simulations are performed on a machine with an Intel(R) Xeon(R) CPU E5-2650 and 128 GB of RAM. The UC simulations use PLEXOS 9.100 [27], a MILP-based electricity market simulation software that the Chilean ISO uses to obtain day-to-day dispatch scheduling with a 168-hour horizon. The mathematical problem formulated by PLEXOS is solved with Gurobi version 9.5.2 [28].

A. DATA

The Chilean ISO website is the source of the nodal load, VRE scheduling, generation scheduling for the reservoirs, and historical operational data of the units [29], [30]. These datasets encompass the actual and planned system operations spanning from January 1, 2019, to December 31, 2022, with an hourly resolution. The full dataset is partitioned into training, validation, and testing sets. Data from 2019 to 2021 are used for training, the first two months of 2022 are used for validation, and the rest of 2022 data are used for testing purposes. Based on the results of the validation dataset, we picked nine thermal units to predict and fix their commitment status when solving the reduced MILP. The selection of those units is based on their relative importance to the total operational cost of the UC problem and the number of binary variables that every unit represents. The subset of nine thermal units whose commitment status is fixed in the reduced MILP problem represents 1.8 GW of power and 129 binary variables.

B. ANNS TUNING

The validation dataset was used for three purposes: (i) to select the generation units whose statuses are to be predicted; (ii) to select the generation units corresponding to each agent of the MADRL scheme; and (iii) to tune the ANN's hyperparameters. In the end, six ANN agents represent the on/off status of the nine units. We grouped some units because they are in nearby areas and tend to behave similarly. Table 1 shows each agent's number of units and capacity.

Some of the hyper-parameters tuned are the number of filters and stacked residual blocks, kernel sizes, and dilation factors. The length of the input (the amount of data to consider from the past) is also an adjustable hyperparameter. Table 2 shows the search space for every parameter and their final values.

TABLE 1. Number of units and capacity per agent.

Agent	Number of units	Capacity (MW)
1	2	57
2	2	762
3	2	120
4	1	255
5	1	370
6	1	248

TABLE 2. Hyper-parameter values for the ANNs.

Hyper-parameter	Search space	Final value
Length input	72, 96, 120, 144, 168	168
Dense	10 - 64	48
Kernel size	3 - 4	3
Filters	16 - 64	16
Blocks	1 - 2	1
Dilations	[1, 2, 4, 8, 16], [1, 3, 9, 27, 81], [1, 3, 6, 12, 24], [1, 5, 7, 14, 28].	[1, 2, 4, 8, 16]

Moreover, an ablation experiment using the validation set is presented in Table 3, exploring different input lengths. The prediction performance is evaluated using metrics such as accuracy, precision, recall, and specificity [31]. These metrics provide insights into the effectiveness of the proposed model in correctly predicting the status of the thermal units. The results of the input lengths tested indicated that the 168-length input yielded the best overall performance. Notably, while specificity and accuracy exhibited relatively consistent performance across different input lengths, both recall and precision demonstrated significant variations. This can be attributed to an imbalance observed in certain power plants between on and off statuses, posing a challenge for the neural network to capture the minority class effectively. Nevertheless, as we have seen with the specificity and accuracy, adopting a strategy of setting zeros in the UC instances leads to high performance. This approach improves the model's effectiveness, reduces the impact of challenges, and enables the network to achieve outstanding performance.

It is important to consider that the inclusion of a large number of power plants in the UC can adversely affect the performance of the MADRL model and computational efficiency. Therefore, it is necessary to investigate the optimal number of power plants to be included in the UC for optimal model performance and computational efficiency.

The training iterations are customized by implementing an early stopping criterion of 50 iterations. We use *ReLU* activations for the neurons and Adam optimizer for training, which has an adaptive learning rate to enhance the convergence of the networks [32]. The average execution time of each training iteration is 6.1 seconds. It is important to consider that a TCN architecture processes sequences much faster than recurrent networks. Fig. 4 shows the evolution of training and validation cumulative reward for the final ANNs.

TABLE 3. Best results for different length inputs.

Length input	Dense	Kernel	Filters	Blocks	Dilations	Recall	Specificity	Precision	Accuracy
72	64	4	28	1	[1,3,6,12,24]	0.6881	0.9427	0.7222	0.9724
96	60	4	56	1	[1,3,9,27,81]	0.7098	0.9411	0.7450	0.9730
120	35	3	50	1	[1,5,7,14,28]	0.6702	0.9370	0.7549	0.9726
144	42	3	30	1	[1,3,6,12,24]	0.6876	0.9445	0.7849	0.9720
168	48	3	16	1	[1,2,4,8,16]	0.7108	0.9469	0.7942	0.9742

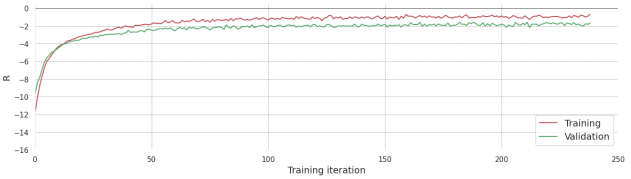

FIGURE 4. Evolution of training and validation reward for the selected ANNs.

V. RESULTS

A. PREDICTION PERFORMANCE OF THE OFFLINE STAGE

As the unit commitment cases in PLEXOS and their results are publicly available on the website of the Chilean ISO [30], we can readily compare the commitment statuses of the selected units predicted by the MADRL model against actual PLEXOS results.

Results are shown in Table 4. For the test set, the obtained results underscore the model's capability to accurately predict periods where the thermal units are off, as evidenced by high accuracy and specificity values. However, it is relevant to note that recall and precision exhibit values around 0.76. Nonetheless, as we mentioned before, adopting a strategy of fixing zeros in the UC instances enhances the model's capability, mitigating the impact of this challenge and allowing the network to excel in performance.

TABLE 4. Prediction performance of the MADRL-based offline stage.

Recall	Specificity	Precision	Accuracy
0.7556	0.9624	0.7614	0.9704

Notice that Table 4 compares predictions made by the MADRL model with simulation results of PLEXOS original solution. The quality of the solutions in terms of the objective function value and computational performance are discussed next.

B. COMPUTATIONAL PERFORMANCE OF THE PROPOSED MODEL

This section compares the normal MILP resolution, which is the industry standard for this type of problem, against solutions obtained with the proposed MADRL-based solution-space reduction approach. It is important to note that the resolution of the original PLEXOS instance, employed by the Chilean Independent System Operator, serves as a pertinent and robust benchmark for our methodology. In practical terms, in the online model the on/off status determined by each ANN-based agent is given to PLEXOS as a structured CSV input file that the software will use to automatically generate the set of corresponding constraints that reduce the solution space. For the computational performance comparison to be fair, we need to rerun the original PLEXOS simulations (i.e. without reducing the solution space) in the same computer where we are running the UC optimization with a reduced solution space. Thus, as the resolution of the unreduced PLEXOS instance is the being the current procedure employed by the Chilean ISO, the resolution of the original PLEXOS instance serves as a benchmark for our proposed methodology. As the solution times may take several hours, our testing is conducted for a limited number of days. Also, to comprehensively represent diverse hydrological conditions within a single year, we conducted tests for 60 randomly selected days from March to December 2022. This approach aims to capture a wide range of scenarios and account for the varying levels of hydroelectric production. Notably, October, November, and December are characterized by increased hydroelectric output due to snowmelt. Remarkably, our proposed method exhibited consistent performance across both rainy months and snowmelt periods, with no significant differences observed. These results suggest that our approach is robust and can provide reliable solutions regardless of the prevailing hydrological conditions.

In the full MILP case, the average solution time is 65.9 minutes. Meanwhile, the average solution time for our proposed framework is only 48.7 minutes. This represents an average reduction time of 26.1%. Of course, the performance may vary depending on the UC instance being solved. Fig. 5 shows boxplots for the MILP time resolution. The interquartile range for the full case falls between 30 and 100 minutes, whereas, for the proposed model, it ranges between 25 and 65 minutes, demonstrating the effectiveness of our method in terms of resolution time. Additionally, we obtained an average speed-up of 1.6, which further confirms the computational efficiency of our approach. The significant reduction in computation times and variance demonstrates our model's efficiency and reliability in delivering consistent performance under diverse operational conditions. This reliability is crucial for power system operators, especially in scenarios requiring quick adaptation to changing hydrological conditions or demand patterns.

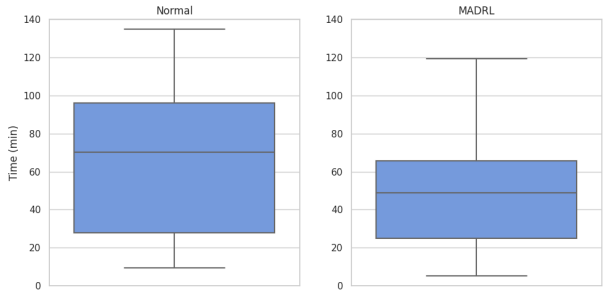

FIGURE 5. Solution time for the MILP-based UC.

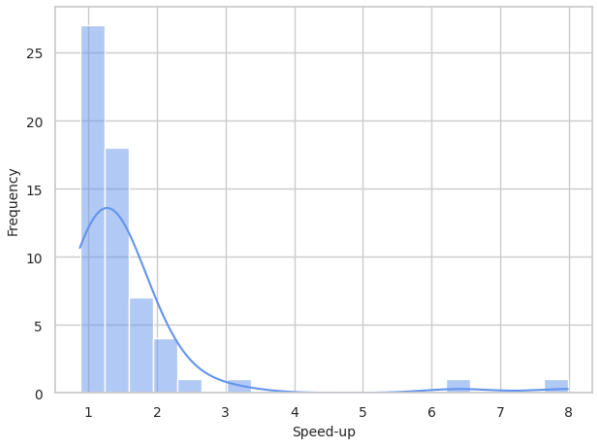

FIGURE 6. Distribution of speed-up.

Fig. 6 shows the distribution of speed-up for our proposed framework. It can be observed that the majority of the tested cases have speed-ups ranging between 1 and 2. The speedups can even reach values between 6 and 8 in certain instances. We conducted a t-test to assess the time reduction achieved by our method. A p-value of 6.46×10^{-3} indicates a statistically significant reduction in time. Also, an F-test revealed that there is a statistically significant reduction in the variance of the solution time (p-value of 4.92×10^{-2}).

The main objective of our work is to achieve a method that would accelerate resolution times while not compromising the solution quality in terms of the total cost. This means that the solution provided by our method needs to fall between the lower and upper bounds set by the original solution's gap. Fig. 7 shows the normalized cost for all test cases. An upper and lower bound, calculated as an average between the different bounds, is included. In most cases, the solution falls within the established margins, and in some cases, more cost-efficient solutions are achieved. Our findings, particularly the operational cost efficiencies and the model's adaptability to different hydrological conditions, align with the industry's broader goals of enhancing sustainability and reliability in power systems. By enabling more efficient unit commitment decisions, our methodology supports the transition towards greater integration of renewable energy sources, contributing to reducing carbon emissions and fostering a more sustainable energy landscape.

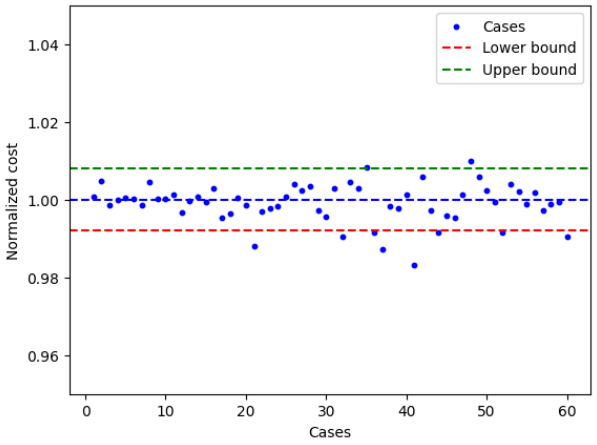

FIGURE 7. Normalized cost.

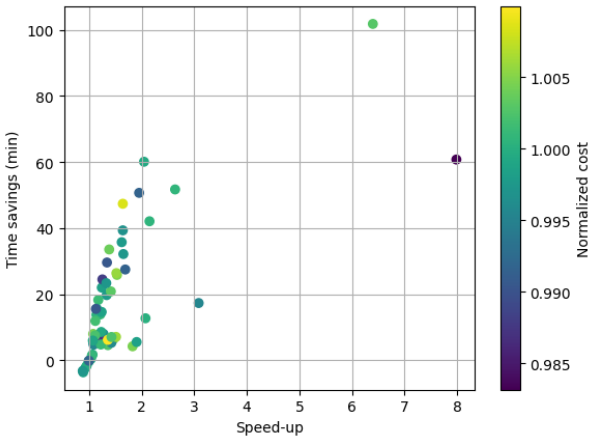

FIGURE 8. Comparison between saving time, speed-up and operational cost.

Fig. 8 compares the speed-up, time savings, and operational cost. Significant time savings are observed when using our model, obtaining, in some cases, more than 40 minutes of reduced execution time. In the few instances where the speed-up is less than one, the time difference does not exceed 3 minutes. Furthermore, these cases have a lower operational cost than the full case. The synergy between time savings and acceptable cost variation vividly underscores the efficacy and feasibility of our innovative methodology tailored to address the unit commitment problem within hydrothermal power systems. This approach holds substantial promise for real-world adoption in the Chilean context, positioning it as a viable and practical solution.

In summary, our results validate the MADRL-based approach as a powerful tool for improving the computational efficiency of the UC problem in hydrothermal systems without compromising on solution quality. The demonstrated computational improvements and cost efficiencies highlight the potential of this methodology for real-world application, offering significant contributions to the fields of machine learning and power system optimization.

VI. CONCLUSION

This article presented an innovative approach that utilizes artificial neural networks to enhance the resolution process of the UC problem in hydrothermal systems. Using ANNs, we accurately determine the thermal units' on/off status, reducing the solution space of the UC problem by employing ANN-based agents and leveraging historical power system operational data. These statuses serve as inputs to solve the UC hydrothermal problem, employing a solution space judiciously reduced for efficiency.

Our study demonstrates a practical application of MADRL in solving the UC problem in hydrothermal power systems and enriches the theoretical landscape of machine learning by showcasing the adaptability and efficacy of MADRL in complex, real-world scenarios. This work bridges the gap between advanced computational techniques and power system optimization, offering new theoretical insights that underscore the potential of machine learning models to enhance decision-making processes in energy systems significantly.

All tests were conducted on actual instances of the Chilean National Electric System's 168-hour ahead generation scheduling processes. Our results demonstrate that the proposed method significantly accelerates the resolution times of practical instances of the unit commitment problem. Furthermore, most evaluated cases indicate that the cost associated with our approach remained within the established error margins, highlighting the method's feasibility and potential for practical application.

A potential limitation arises when the system faces atypical hydrological conditions. While the state of the reservoirs is used as input to account for hydrology, accurately predicting this variable is challenging. This can pose difficulties for the model, especially under extreme hydrological scenarios. Moreover, the current phase of energy transition, marked by the phasing out of thermal power plants and the integration of new technologies, demands regular updates to the model.

In practical terms, our framework is engineered to streamline the UC problem resolution. The retraining of offline agents depends critically on access to current hydrological data, a vital factor in hydrothermal systems. It's essential to acknowledge that power systems are changing, with significant transitions like the decommissioning of thermal plants and the adoption of innovative technologies. Such changes require periodic retraining of the model, ensuring its continuous alignment with the evolving dynamics of the system.

Avenues for further research include exploring new strategies for fixing binary variables, which could lead to further efficiency in solving the UC problem. Also, it is worth mentioning that including too many power plants impaired the generalization of the ANN, resulting in poorer performance. Thus, systematically exploring the optimal number of agents and power plants associated with each one could improve the performance. Moreover, incorporating additional operational constraints during the neural network training could enhance the accuracy of predictions and the method's applicability in real-world contexts.

REFERENCES

[1] Y. Yang and L. Wu, "Machine learning approaches to the unit commitment problem: Current trends, emerging challenges, and new strategies," *Electr. J.*, vol. 34, no. 1, Jan. 2021, Art. no. 106889.

[2] C. Wang and S. M. Shahidehpour, "Effects of ramp-rate limits on unit commitment and economic dispatch," *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 1341–1350, Mar. 1993.

[3] M. P. Walsh and M. J. O'Malley, "Augmented Hopfield network for unit commitment and economic dispatch," *IEEE Trans. Power Syst.*, vol. 12, no. 4, pp. 1765–1774, May 1997.

[4] R. Nayak and J. D. Sharma, "A hybrid neural network and simulated annealing approach to the unit commitment problem," *Comput. Electr. Eng.*, vol. 26, no. 6, pp. 461–477, Aug. 2000.

[5] Y. Zhou, Q. Zhai, L. Wu, and M. Shahidehpour, "A data-driven variable reduction approach for transmission-constrained unit commitment of large-scale systems," *J. Modern Power Syst. Clean Energy*, vol. 11, no. 1, pp. 254–266, Jan. 2023.

[6] A. M. Alvarez, Q. Louveaux, and L. Wehenkel, "A machine learning-based approximation of strong branching," *INFORMS J. Comput.*, vol. 29, no. 1, pp. 185–195, Jan. 2017.

[7] Á. S. Xavier, F. Qiu, and S. Ahmed, "Learning to solve large-scale security-constrained unit commitment problems," *INFORMS J. Comput.*, vol. 33, no. 2, pp. 739–756, Oct. 2020.

[8] A. V. Ramesh and X. Li, "Feasibility layer aided machine learning approach for day-ahead operations," *IEEE Trans. Power Syst.*, vol. 39, no. 1, pp. 1582–1593, 2024.

[9] Z. Lin, Y. Chen, J. Yang, C. Ma, H. Liu, L. Liu, L. Li, and Y. Li, "Accelerating transmission-constrained unit commitment via a data-driven learning framework," *Frontiers Energy Res.*, vol. 10, Jan. 2023, Art. no. 1012781.

[10] A. Venkatesh Ramesh and X. Li, "Spatio-temporal deep learning-assisted reduced security-constrained unit commitment," 2023, *arXiv:2306.01570*.

[11] Y. Tang, Z. Pan, X. Hu, W. Pedrycz, and R. Chen, "Knowledge-induced multiple kernel fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14838–14855, 2023.

[12] Z. Yan and Y. Xu, "Real-time optimal power flow: A Lagrangian based deep reinforcement learning approach," *IEEE Trans. Power Syst.*, vol. 35, no. 4, pp. 3270–3273, Jul. 2020.

[13] T. Ochoa, E. Gil, A. Angulo, and C. Valle, "Multi-agent deep reinforcement learning for efficient multi-timescale bidding of a hybrid power plant in day-ahead and real-time markets," *Appl. Energy*, vol. 317, Jul. 2022, Art. no. 119067.

[14] B. Knueven, J. Ostrowski, and J.-P. Watson, "On mixed-integer programming formulations for the unit commitment problem," *INFORMS J. Comput.*, vol. 32, no. 4, pp. 857–876, Jun. 2020.

[15] E. Gil, J. Bustos, and H. Rudnick, "Short-term hydrothermal generation scheduling model using a genetic algorithm," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1256–1264, Nov. 2003.

[16] M. V. F. Pereira and L. M. V. G. Pinto, "Application of decomposition techniques to the mid-and short-term scheduling of hydrothermal systems," *IEEE Power Eng. Rev.*, vol. PER-3, no. 11, pp. 37–38, Nov. 1983.

[17] P. Hernandez-Leal, B. Kartal, and M. E. Taylor, "A survey and critique of multiagent deep reinforcement learning," *Auto. Agents Multi-Agent Syst.*, vol. 33, no. 6, pp. 750–797, Nov. 2019.

[18] S. Gronauer and K. Diepold, "Multi-agent deep reinforcement learning: A survey," *Artif. Intell. Rev.*, vol. 55, no. 2, pp. 895–943, Feb. 2022.

[19] W. Du and S. Ding, "A survey on multi-agent deep reinforcement learning: From the perspective of challenges and applications," *Artif. Intell. Rev.*, vol. 54, no. 5, pp. 3215–3238, Jun. 2021.

[20] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022.

[21] T. Standley, A. Zamir, D. Chen, L. Guibas, J. Malik, and S. Savarese, "Which tasks should be learned together in multi-task learning?" in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 9120–9132.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[23] S. Bai, J. Zico Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
[25] A. N. Tarekegn, M. Giacobini, and K. Michalak, "A review of methods for imbalanced multi-label classification," *Pattern Recognit.*, vol. 118, Oct. 2021, Art. no. 107965.
[26] Generadoras de Chile. *Capacidad De Generación En Chile*. Accessed: Dec. 2022. [Online]. Available: http://generadoras.cl/generacion-electrica-en-chile
[27] Energy Examplar. *PLEXOS for Power Systems: Power Market Simulation and Analysis Software*. Accessed: Dec. 2022. [Online]. Available: https://www.energyexemplar.com/plexos
[28] Gurobi Optimization, LLC. (2023). *Gurobi Optimizer Reference Manual*. [Online]. Available: https://www.gurobi.com
[29] Coordinador Eléctrico Nacional. *Operación Real*. Accessed: Dec. 2022. [Online]. Available: https://www.coordinador.cl/operacion/graficos/operacion-real/
[30] Coordinador Eléctrico Nacional. *Generación Programada*. Accessed: Dec. 2022. [Online]. Available: https://www.coordinador.cl/operacion/graficos/operacion-programada/generacion-programada/
[31] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manage. Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015.
[32] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

PHILIP GUERRA (Student Member, IEEE) was born in Illapel, Chile. He received the B.Sc. and M.Sc. degrees in electrical engineering from Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2020 and 2023, respectively.

From 2023 to February 2024, he was a Planning Engineer with Coordinador Eléctrico Nacional, Santiago, Chile. His research interests include power system planning and operation and using artificial intelligence methods in power systems.

ESTEBAN GIL (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Universidad Técnica Federico Santa María (UTFSM), Valparaíso, Chile, in 1997 and 2001, respectively, and the M.Sc. degree in statistics and the Ph.D. degree in electrical engineering from Iowa State University, in 2006 and 2007, respectively.

He has more than 20 years of experience working as a Consultant in Australia and Chile modeling electricity markets in Oceania, Asia, Africa, and South America. He is currently a Professor with UTFSM and a Researcher with the Advanced Center for Electrical and Electronic Engineering (AC3E).

VÍCTOR H. HINOJOSA (Member, IEEE) was born in Quito, Ecuador. He received the B.Sc. degree in electrical engineering from Escuela Politécnica Nacional (EPN), Quito, in 2000, and the Ph.D. degree in electrical engineering from Universidad Nacional de San Juan, Argentina, in 2007.

He is currently a Professor with the Department of Electrical Engineering, Universidad Técnica Federico Santa María, Valparaíso, Chile. His research interests include power system dynamics, modeling and optimization, planning and operation of power systems, and integration of renewable energies.