The emerging threat of artificial intelligence on competition in liberalized electricity markets: A deep Q-network approach

Danial Esmaeili Aliabadi[a,*], Katrina Chan[a]

[a]*Helmholtz Centre for Environmental Research - UFZ, Permoserstraße 15, 04318, Leipzig, Germany.*

Abstract

According to Sustainable Development Goals (SDGs), societies should have access to affordable, reliable, and sustainable energy. Liberalized electricity markets have been established to provide affordable electricity for end-users through advertising competition. Although these new markets are designed to serve competition, there are recorded incidents where participants abused their market power and disrupted the competition through collusion. Unfortunately, modern autonomous pricing algorithms may further assist myopic players to discover collusive strategies with a minimum amount of sensitive information. Therefore, in this study, we investigate the impact of emerging learning algorithms on the bidding strategies of Power Generating Companies (GenCos) and compare their performance against game-theoretic expectations. A novel deep Q-network (DQN) model is developed, by which GenCos determine the bidding strategies to maximize average long-term payoffs in a day-ahead market. The presented DQN model assumes that GenCos have no information regarding the rivals' true generation costs and profits. To the best of the authors' knowledge, this is the first study that thoroughly investigates players' behavior utilizing a modern DQN model and compares its results with equilibria of the non-cooperative single-stage and infinitely-repeated games in the context of electricity markets. The outcomes articulate that GenCos equipped with advanced learning models may be able to collude unintentionally while trying to ameliorate long-term profits. Moreover, GenCos that employ the presented DQN model could discover and sustain more profitable (e.g., collusive) strategies vis-à-vis a conventional Q-learning method. Collusive strategies can lead to exorbitant electric bills for end-users, which is one of the influential factors in energy poverty. Thus, policymakers and market designers should be vigilant regarding the combined effect of information disclosure and autonomous pricing, as new models exploit information more effectively.

Keywords: Collusion, deep Q-network, day-ahead electricity market, Nash equilibrium

1. Introduction

The 7th United Nations' Sustainable Development Goal (SDG7) invites societies to provide affordable, reliable, and sustainable energy for everyone. While access to clean and reliable energy is a major concern in many developing countries [1], the developed world has been focused on energy affordability [2]. Due to the unfavorable economics of electricity storage technologies and the constant need for balancing operating resources and loads in real-time, the electricity industry expanded as vertically integrated monopolies in the past, which subsequently increased the operating costs [3, 4]. The high electricity price is one of the influential factors causing energy poverty in societies [5]. To ensure electricity affordability, governments have pursued liberalization (i.e., deregulation) that aims to maximize social welfare through promoting competition among self-interested participants [6]. Although market designers expect to witness full competition, it is demonstrated that some electricity markets act more like oligopolies for the following reasons [7]:

- Limited number of generators as a result of high capital investment.
- Network congestion that prevents generators from dispatching power to inaccessible consumers.
- Transmission losses that hinder producers in serving remote consumers.

Oligopolistic markets may incubate collusion that harms open competition among participants. While explicit collusion in electricity markets is prohibited, tacit collusion may still exist in the absence of formal contracts [8]. Heim and Götz [9] study the rising price of reserve power in the German market. The authors conclude that the seemingly collusive behavior is due to the repetitive auctions with the pay-as-bid pricing mechanism. Similarly, many studies believe that participants engage in collusive behavior in the electricity markets of the UK [10, 11], Spain [12], and California [13]. To achieve a perfectly competitive market, collusion (of any kind) should be eliminated, but it is not a straightforward task for regulators to detect tacit collusion [14, 15].

To make matters worse, antitrust agencies are worried that the autonomous pricing algorithms, often used by suppliers, may learn to collude unintentionally [16, 17]. Since the advent of Deep Reinforcement Learning (DRL), algorithmic pric-

*Corresponding Author; Tel.: +49-341-2434-581

Email addresses: danial.esmaeili@ufz.de (Danial Esmaeili Aliabadi), katrina.chan@ufz.de (Katrina Chan)

ing has attracted more attention [18]. Nowadays, these algorithms are common in many markets; for instance, according to Chen et al. [19], 500 vendors out of 1,641 on Amazon Marketplace benefited from automated pricing algorithms. An efficacious pricing algorithm can provide the power producers in the market an unfair cutting edge to increase the electricity price and damage consumers financially; therefore, it is worth evaluating the impact of algorithmic pricing under the current market structure.

In this study, we aim to create a state-of-the-art learning model based on Deep Neural Networks (DNN) that assists generic players, to raise and sustain their incomes without using confidential information related to employed technologies (e.g., the unit generation cost) while also taking transmission network constraints into account. The outcomes are then investigated to assess the possibility of players unintentionally engaging in collusive behavior. For this purpose, the results of the offered learning model are compared with game-theoretic expectations [20] and an extended version of the conventional Q-learning algorithm [21, 22] in a setting with collusive equilibria. Although DNN models are applied to a wide variety of problems, to the best of the authors' knowledge, this manuscript is the first to analyze the capacity of the aforementioned models in sustaining collusive behavior in liberalized electricity markets based on a complex case with multiple Nash and collusive equilibria.

The remainder of the manuscript is organized as follows: In Section 2, we present the literature focusing on various modeling techniques and their applications in liberalized electricity markets. Section 3 defines the problem and introduces two learning algorithms for players. The hyper-parameters of the proposed algorithms are also adjusted in this section. In Section 4, the developed learning algorithms are applied to a case study with collusive strategies. The outcomes are contrasted with equilibria of the non-cooperative single-stage and infinitely-repeated games to understand the impact of advanced learning algorithms on strategic decision-making. Section 5 discusses the practical challenges that can hinder the application of these techniques in real-world markets. Finally, Section 6 concludes.

2. Literature review

2.1. Optimization models

From the literature, one can name two overarching market modeling trends: optimization (including equilibrium) and simulation. Optimization models, which are employed extensively in formulating the strategic behavior of profit-driven power Generating Companies (GenCos), often require knowledge about rivals' confidential information and the market clearing mechanism [23, 24]. Aliabadi et al. [20] formulate collusion among participants as a bi-level model and provide sufficiency conditions for its existence using the folk theorem. The authors also propose a heuristic algorithm to compute a collusive state given that all Nash equilibria are known. Ebadi Torkayesh [25] proposes two new formulations based on the developed bi-level model in Aliabadi et al. [20]. For bi-level models, in particular, the lower level should be free of any binary variables, since it is replaced with the Karush Kuhn-Tucker (KKT) optimality conditions [26]. As such, solving the resulting optimization model could present distorted outcomes, as it would lack variables that capture real-world behaviors including shut-down and start-up [27]. Finally, as mentioned earlier, the deterministic optimization models demand strict assumptions such as the perfect knowledge of the market's and competitors' operating parameters, which can render the results of optimization models unreliable. [28].

2.2. Simulation models

As an alternative to optimization (and equilibrium), simulation models can be utilized when underlying problems are intractable through analytical methods [29]. Typically, researchers rely on agent-based simulation models in decentralized electricity markets, since it provides sufficient flexibility to investigate the impact of learning on GenCos' strategic behavior [30]. At the forefront of imitating human-like intelligence are model-free Reinforcement Learning (RL) algorithms [31]: agents learn the optimal set of actions (i.e., optimal policy) with respect to each state, solely by interacting with the environment.

Previous studies have utilized RL to analyze competition in electricity markets. Staudt et al. [32] studied the number of suppliers needed to secure competition in a local electricity market and investigate the impact of tacit collusion through signaling. The authors infer that peak capacity has to be provided by multiple suppliers in order to guarantee competitive prices. Emami et al. [33] propose a simulation-based method to assess possible coalitions in a wholesale electricity market, in which GenCos can negotiate among themselves. Poplavskaya et al. [34, 35] employ agent-based simulation models with RL-equipped agents to investigate actors' behavior in a common standalone balancing energy market, which is scheduled to be put into practice in 2022. Namalomba et al. [36] model the double-sided auction with elastic demand in a centralized electricity market using a bi-level model and Q-learning algorithm and compare outcomes with Nash equilibrium. The authors conclude that energy transaction prices can be reduced with the participation of flexible consumers. Jia et al. [37] propose a continuous action RL algorithm to help GenCos in making more profitable decisions in environments with limited access to market information.

In an oligopolistic electricity market, Aliabadi et al. [21] show that agents with a time-dependent Q-learning algorithm can converge to either Nash equilibria or strategies with the same profit gains under most parameters combinations. On the other hand, Klein [38] show that RL-equipped agents can find collusive equilibria in a simple duopoly setting. Calvano et al. [16] examine an oligopoly market and demonstrate that increasing the number of agents can weaken the performance of RL and decrease the profit gain. In this study, we are interested to know whether a cutting-edge learning algorithm can attain collusion in more sophisticated environments.

In spite of their success in various fields – including operations research, decision, and control theories – RL methods (e.g., Q-learning) suffer from two major drawbacks: a lack of theoretical proof to assure convergence to optimality [38], and

the curse of dimensionality [39]. As the state space expands, the required memory to store transitions grows exponentially with it. To circumvent the dimensionality curse, *Roth-Erev* learning [40] has been developed, which is a streamlined version of RL when a limited number of pure strategies are played by agents. However, *Roth-Erev*-equipped agents are unable to learn consistent behaviors in complex games, such as the sequential bargaining game [41].

2.3. Deep Neural Networks models

A more recent trend to address the dimensionality challenge is to estimate the optimal action-selection policy using DNN. Artificial Neural Networks (ANNs) have been used in various fields [42, 43] since the 1950s; nevertheless, the combination of ANNs and RL algorithms together with the ever-increasing computational power and the availability of big data attracted researchers' attention in the field of Artificial Intelligence (AI) [44]. The Deep Q-Network (DQN) models train a DNN structure using supervised learning and RL. Due to versatility, DQN models are employed in many applications, ranging from the agents' decision-making in classic console games to the more sophisticated board game *Go* [18, 45].

In 2019, a DQN model was developed for the first time to optimize GenCos' bidding strategies in liberalized electricity markets [27]. The same group extended their model in 2020 and compared their results with the conventional Q-learning and bi-level models [24]. In both studies, the authors conclude that their DQN models outperform conventional Q-learning models; however, they did not examine agents' behavior using the proposed DQN model in the presence of collusion. In 2020, Liang et al. [46] proposed a Deep Deterministic Policy Gradient (DDPG) algorithm based on the actor-critic framework and showed that two GenCos exploiting the proposed DDPG algorithm converge to the Nash equilibrium of a complete information game. Through simulation, they also exhibit that GoCos can increase the market cleared price when the discount factor is sufficiently close to one; however, the divergence from the Nash equilibria might not be a valid measure to signal tacit collusion. For instance, the authors inferred that GenCos equipped with a Q-learning algorithm converge to collusion by playing a suboptimal strategy, while GenCos with DDPG converge to the Nash equilibrium. This counter-intuitive result cannot be generalized using the folk theorem, as shown by Aliabadi et al. [21]. Therefore, in this study, we thoroughly investigate the impact of strategic bidding behavior on the GenCos' payoff profiles and compare outcomes with multiple analytically calculated solutions (i.e., discount factor = 0 and 1).

Recently, Razmi et al. [15] employ supervised learning algorithms to detect collusion in day-ahead markets. This algorithm can be used by independent system operators in markets with limited dynamism. Likewise, Pan et al. [47] and Velloso and Van Hentenryck [48] tackle technical challenges and suggest DNN models, by which the regulator can estimate a near-optimal solution to a large-scale optimal power flow problem in a reasonable time. Guo et al. [28] propose a data-driven recognition system for a bidding objective function using deep inverse reinforcement learning and verify their results using DQN. The results demonstrate that an advanced algorithm can extract sensitive information about participants based on their bidding behavior.

Renewable GenCos and innovative concepts have endeavored to employ DRL according to their conditions. For instance, in 2021, Lehna et al. [49] developed a DRL algorithm for wind park operators, by which the producers can increase the total net profit on the intraday electricity market. The proposed model outperforms several baselines in the continuous German intraday market with one-minute temporal resolution. As mentioned by Scholz et al. [50] and Nolting et al. [51], solving energy system models and electricity market models with such a high temporal resolution via optimization is computationally costly; nonetheless, its importance is increasing due to the growing shares of intermittent renewable energy sources. As shown by Owolabi et al. [52] in the US market, a high level of variable renewable energy can lower electricity prices, while affecting the price volatility non-linearly. Finally, Löschenbrand [53] model the competition between virtual power plants using deep learning.

To expose the gap in the literature, Figure 1 compares most-related studies to the current paper with respect to multiple criteria. As illustrated, most studies are at the intersection of two or three criteria, while there are only two studies at the center.

3. Methodology

3.1. Market clearing mechanism

In this paper, the strategic bidding problem of GenCos on a day-ahead market is considered, taking network constraints into account. A typical electric grid is made of interconnected nodes (i.e., regions), which function independently. In each node, the produced power by GenCos is consumed by demand centers, and the excess power flows to the connected nodes through transmission lines. Due to physical limitations, transmission lines are unable to dispatch electricity above a certain threshold. A power network is called "congested" when a thoroughly loaded transmission line reaches its maximum capacity and cannot accommodate further dispatch. The Independent System Operator (ISO) manages network congestion by penalizing electricity consumption at congested nodes using the Locational Marginal Pricing[1] (LMP) scheme [64].

To manage the day-ahead market, the ISO conducts a series of auctions every day, in which GenCos submit their bid prices ($b_i^t \in \mathcal{B}_i$) and feasible production capacities (P_i^L and P_i^H) for each hour of the next day ($t \in \{1, \ldots, 24\}$). $\mathcal{B}_i$ is the set of all feasible bid values for GenCo-i. It is common in many markets that ISO sets an upper limit for the submitted bid prices (i.e., $b_i^t \leq b_i^{max}$) to prevent unreasonable electricity prices [65]. Subsequently, the ISO solves an optimal power flow problem concerning submitted bids such that social welfare is maximized at each hour. In this manuscript, the Direct Current Optimal Power Flow (DCOPF) problem [66] is adopted as it is

[1]The LMP scheme (i.e., nodal pricing) is practiced in the US electricity markets while Europe uses slightly different zonal pricing. [63]

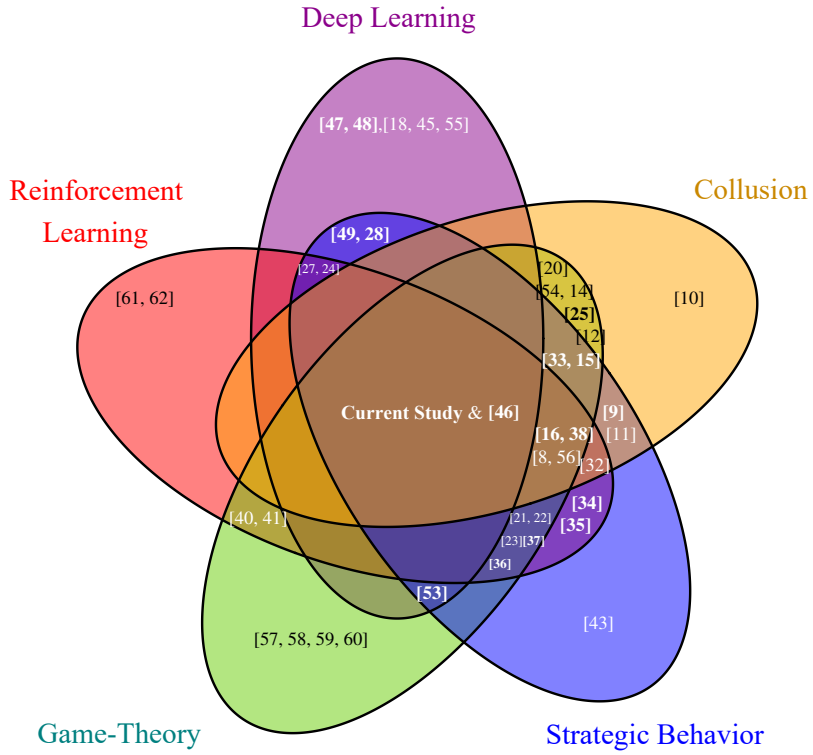


Figure 1: Reviewed studies categorized into five groups. Most studies are at the intersection of multiple criteria; however, there are only two studies at the center: the current study and [46]. Bold numbers represent recent studies (i.e., 2020 and afterward).

employed extensively in power systems operation (e.g., Power World, GridView, MAPS, and Promod) [67] and is a linear programming model. The optimal solution of the DCOPF problem at hour t determines the electricity price (λ_i^t) and voltage angle (θ_i^t) at each node, and GenCos' production level (P_i^t). The DCOPF formulation is provided in Appendix A.

After clearing the market by the ISO, GenCos can calculate their payoffs at each specific hour as $r_i^t = P_i^t(\lambda_i^t - c_i)$, where the electricity generation cost of GenCo-i is captured by c_i. It is quite realistic to assume GenCos conceal their payoffs from rivals as it may reveal confidential information regarding their business [28].

3.2. Assumptions and limitations

In this manuscript, we introduce an unexplored issue with major impacts on future market designs and policies. To establish a theoretical base for our discussions, the equilibria solutions (e.g., the Nash equilibria) are calculated. Computing game-theoretic expectations comes at a cost: The proposed setup for this study considers a simplified day-ahead market; however, real-world electricity markets are complex and include features like intraday trading, auxiliary services, and block bids [68, 69]. Unfortunately, finding analytical solutions for such complex problems is hopeless [58]. Therefore, we rely on assumptions that might be undesirable from practitioners' points of view.

Similar to [21, 22], the following assumptions are considered in the presented model:

- The ISO considers the network structure and clears the day-ahead market using the DCOPF problem.
- Generic GenCos are taken into account; thus, GenCos can utilize various technologies (e.g., biogas power plants, wind turbines).
- The demand is assumed to be inelastic to obtain theoretical solutions for the collusive strategies. In this manuscript, theoretical solutions are compared with the simulation results.
- To ease modeling, small players (i.e., GenCos and demand centers) in each node are aggregated; therefore, aggregated GenCos are assumed to be influential players, which means they can affect rivals' strategic behavior. This assumption is not disruptive because of the oligopolistic nature of electricity markets [7, 70]. For instance, only ten companies own over half of the total power production capacity in Turkey[2]. Furthermore, more complex networks with multiple GenCos at each node can still be transformed to an equivalent network with a single GenCo per node using dummy nodes and unbounded transmission

[2]Check the list of companies at `https://www.enerjiatlasi.com/firma/`; accessed on 07 March 2022.

lines. Figure 2 depicts the transformation of the second node with two GenCos in Network-A to an equivalent network (i.e., Network-B) with a single GenCo per node.

3.3. Collusive strategy

As mentioned earlier, a feature of any oligopolistic market is the likelihood of participants engaging in collusion. Collusive strategies can lead to exorbitant electric bills for end-users by damaging consumer surplus in favor of producer surplus, thereby causing energy poverty in societies [5].

Table 1 displays a simplified case, wherein GenCos' payoffs (r_1^t, r_2^t) are displayed in Euro (€) with respect to various bid values. The numbers in Table 1 are excerpted from the first case study in Aliabadi et al. [20]. According to Table 1, (b_1^t = €20/MWh, b_2^t = €30/MWh) is the Nash equilibrium of the single-stage game (i.e., $t \in \{1\}$), as no GenCos can get a better payoff by deviating from the Nash strategy given the other player keeps bidding the same price; nonetheless, there is another strategy (b_1^t = €30/MWh, b_2^t = €40/MWh), the so-called collusive strategy, in which both GenCos can obtain higher payoffs. Reviewing the underlying network [20, Figure 1] reveals that GenCo-1 and GenCo-2 have market power, and GenCo-5 can only provide electricity to the domestic load at the fifth node. Also, neither of these two GenCos can fulfill the demand thoroughly without the other GenCo.

Although the collusive strategy serves both GenCos, it is considered unstable in a single-stage game since GenCo-2 can benefit far more by deviating the collusive strategy, i.e., (b_1^t = €30/MWh, b_2^t = €20/MWh). What prevents GenCo-2 from doing so is the response of GenCo-1 in the forthcoming hours, which can move the game to the Nash equilibrium and damage the long-term profit of both GenCos in infinitely repeated games (i.e., $t \in \{1, \ldots, \infty\}$): GenCo-2's deviation from the collusive strategy (by offering €20/MWh) forces GenCo-1 to play €20 per MWh instead of €30/MWh. In the next hour, GenCo-2 has to offer €30/MWh. Therefore, both GenCos fail by playing a less profitable strategy for the reminder of the time horizon. In the context of game theory, this strategy is called the grim-trigger strategy [71].

Based on the UK's competition market authority report [72], algorithmic pricing may help to improve the stability of collusion by allowing cartel members to identify deviations from the negotiated bid prices more rapidly.

Table 1: Payoff profile of GenCo-1 and GenCo-2. The arrows display the transformation of offers in subsequent hours when GenCo-2 deviates from the collusive strategy. Source: adopted from [20, Table 4].

$B_1 \backslash B_2$	**€20/MWh**	**€30/MWh**	**€40/MWh**	**€50/MWh**
€20/MWh	(857,0) →	(3428, 785)	(6000, 0)	(6000, 0)
€30/MWh	(416, 2500) ←	(3428, 785)	**(6000, 1571)**	(6000, 0)
€40/MWh	(0, 6000)	(0, 6000)	(0, 6000)	(5000, 0)
€50/MWh	(0, 6000)	(0, 6000)	(0, 6000)	(0, 7500)

In this manuscript, we adopt terminology and definitions similar to that which is available in Aliabadi et al. [20] for the Strong Collusive Equilibrium (SCE) and the most collusive state (i.e., SCE*). To summarize, a tuple of bid prices is SCE, if the corresponding payoffs for all GenCos are higher than those under all Nash equilibria ($r_i^{SCE} > r_i^N$). Also, we call a tuple of bid prices SCE* when the minimum payoff is greater than or equal to the minimum payoff for all GenCos in all other SCEs.

Although modeling collusion through optimization and game theory is mathematically elegant, it often requires strict assumptions such as perfect knowledge; however, as practitioners confirm, GenCos have imperfect knowledge about rivals' payoffs in the real world [28]. To address this problem, researchers rely on the simulation of learning agents, a process we also employ in this study.

3.4. Learning

In this manuscript, two learning mechanisms are discussed in detail. The first section is devoted to a simple Q-learning method with time-dependent parameters, which has been employed in Aliabadi et al. [21]. GenCos that benefit from this Q-learning model exploit their past experiences alone. The next section discusses the proposed DQN method. Although GenCos have no information regarding the dispatched power and the generation cost of rivals, the submitted bids to the ISO are assumed to be common knowledge in the proposed DQN model. The outcomes of the mentioned learning methods will be contrasted.

We choose to employ a DQN model over a DDPG model [55] as the DDPG models are especially advantageous when dealing with the continuous action domain; nonetheless, the action space of this study is discrete in nature (i.e., the submitted bid prices have one-cent resolution).

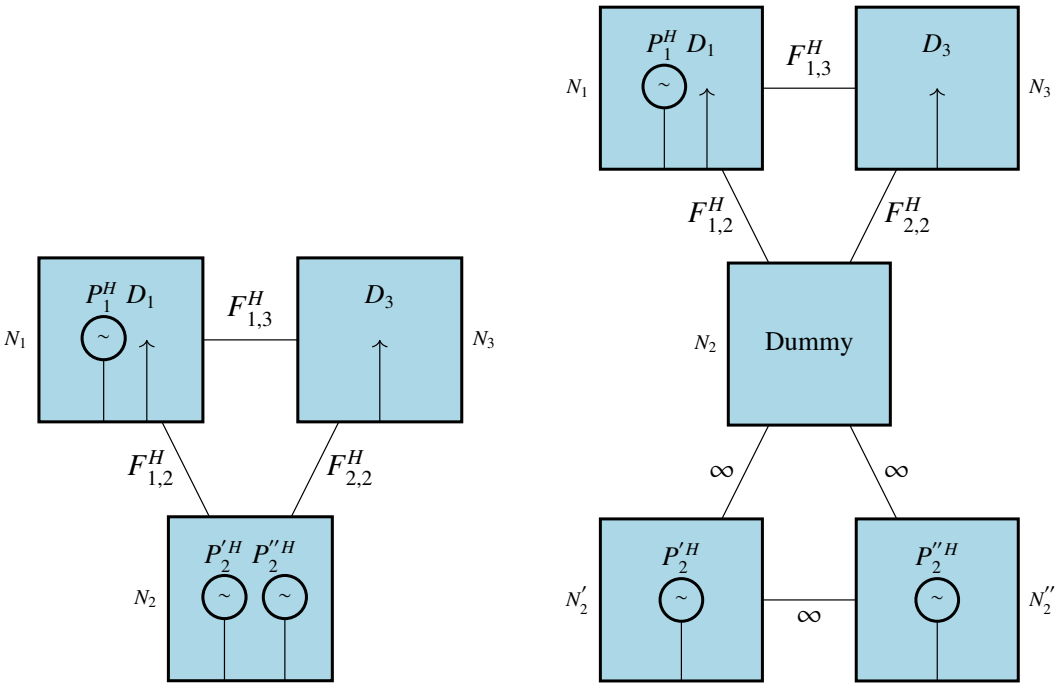
3.4.1. Q-learning with decay

For each hour, agents submit their bid prices to the ISO in order to satisfy the demand. The ISO determines the winning bids and LMPs, taking the transmission network structure into account. For this algorithm, GenCos calculate the profit corresponding to the submitted bid prices, assuming that they have no information regarding the submitted bids by rivals. Consequently, the optimal action of GenCos can vary based on rivals' responses.

To capture the dynamics of such markets, players should associate uneven significance to the information, based on accumulated knowledge. Thereby, the following time-dependent parameters are introduced:

- Recency rate (α_i^t) determines the importance of the recent outcomes for ith GenCo at iteration t. The value of α_i^t is expected to decline as GenCo-i collects information.
- Exploration parameter (ϵ_i^t) adjusts the exploration rate versus exploitation. As GenCo-i becomes mature, it tends to rely more on collected information than searching for undiscovered solutions.

GenCo-i chooses a bid price randomly with the probability ϵ_i^t, whereas the best-known bid, $b_i^* = \underset{b_{ij} \in B_i}{arg\,max}\{Q_{ij}^t\}$, with the probability $1 - \epsilon_i^t$. In the literature, this mechanism is called



(a) Network-A (b) Network-B

Figure 2: Transforming Network-A with multiple GenCos in the second node to Network-B with a single GenCo per node.

the ϵ-greedy action selection rule [73]. Contrary to generic RL algorithms, ϵ_i^t decreases linearly over time to a value near zero, i.e., $\epsilon_i^t = max\{0.001, \frac{8t(\epsilon_i^0-1)}{max_t}+\epsilon_i^0\}$, as GenCo-$i$ explores the state-action space sufficiently.

Furthermore, at each iteration, $t \in \{1, \ldots, max_t\}$, GenCo-$i$ updates the Q-value (Q_{ij}^t) corresponding to each bid price ($b_{ij} \in \mathcal{B}_i$) based on modified α_i^t and the realized payoff (r_{ij}) as described in Eq. (1).

$$\begin{aligned} \alpha_i^t &= \alpha_i^0 - (0.9t/max_t)\alpha_i^0 \\ Q_{ij}^t &= (1-\alpha_i^t)Q_{ij}^{t-1} + \alpha_i^t r_{ij} \end{aligned} \tag{1}$$

3.4.2. Deep Q-Networks approach

In this section, the detail of the proposed DQN model is described, by which GenCos enhance their understandings of the environment and optimize their actions accordingly. The critical elements of the proposed model are as follows:

- **Environment:** The platform whereby ISO clears the market and determines agents' rewards.
- **Agents:** Myopic GenCos that desire to increase their long-term rewards through learning.
- **State:** vector s_i^t encapsulates the state of the system for GenCo-i at time t. In our setting, s_i^t consists of the submitted bid prices by all GenCos at time t in addition to private information related to GenCo-i, such as c_i and P_i^t.
- **Action:** The response of GenCo-i to improve its reward, based on an observed state (i.e., $b_i^t \in \mathcal{B}_i$).
- **Reward:** the obtained payoff of GenCo-i, r_i^t, based on assigned power and cleared price after submitting a bid price.

The overall workflow of the proposed DQN model is depicted in Figure 3. Agents submit random bids at the beginning of the time horizon and store results until the number of records in their replay memory ($\mathcal{M}_i$) exceeds a minimum level. Then, GenCo-i chooses a batch of experiences from memory using the Last-In, First-Out (LIFO) scheme[3]. The LIFO scheme is used to prioritize and capture recent interactions among players. The selected experiences $\{s_i^t, b_i^t, r_i^t, s_i^{t+1}\}$ are normalized and fed into a feed-forward multi-layer neural network to predict the expected reward for the submitted bid price b_i^t using Eq.(2), which is a modified version of the Bellman equation by [74].

$$\begin{aligned} Q_i^{t+1}(s_i^t, b_i^t|\vec{w}_i) = (1-\alpha_i^t)\, Q_i^t(s_i^t, b_i^t|\vec{w}_i) + \\ \alpha_i^t\,(r_i^t + \gamma\, \mathbb{E}[max_{b_i^{t+1}}\{Q_i^t(s_i^{t+1}, b_i^{t+1}|\vec{w}_i)\}] \end{aligned} \tag{2}$$

$$\alpha_i^t = \alpha_i^0\, e^{-0.1\left(|s_i \in \mathcal{M}_i : s_i = s_i^t| - 1\right)} \tag{3}$$

In Eq.(2), the discount factor ($\gamma \in (0,1)$) presents GenCos' perceived significance of future rewards compared to immediate payoff. According to Eq.(2), the expected future reward, $\mathbb{E}[max_{b_i^{t+1}}\{Q_i^t(s_i^{t+1}, b_i^{t+1}|\vec{w}_i)\}]$, is calculated since s_i^{t+1} is established based on the collective actions of all GenCos (b_i^t, $\forall i \in \mathcal{I}$), and not a GenCo solely. As is evident, the Markov property

[3] Our approach is different from Ye et al. [27], which uses the first-in, first-out scheme

does not hold, considering the action space of each GenCo at the beginning of the simulation, i.e., $p(s_i^{t+1}|s_i^t, b_i^t) \neq p(s_i^{t+1}|s_i^1, b_i^1, s_i^2, b_i^2, \ldots, s_i^t, b_i^t)$; however, this property may hold if all GenCos act optimally and choose the best bid, b_i^*, corresponding to a given state at time t. Thus, the Markov property asymptotically holds true if the learning process converges.

Eq.(3) reduces the α_i^t value from an initial level of α_i^0 based on the number of recorded identical s_i^t entries in $\mathcal{M}_i$. Hence, when state s_i^t appears more frequently, $Q_i^t(s_i^t, b_i^t|\vec{w}_i)$ converges to a fixed function, i.e., $Q_i^*(s_i^t, b_i^t|\vec{w}_i)$, as the solution space is being explored sufficiently [61]. To improve the network stability during the learning process, the weight vector ($\vec{w}_i$) of the target network is synchronized periodically (i.e., every 2500 iterations). In theory, this technique assists smoother convergence by preventing instantaneous oscillations while accelerating the process by not training the target network separately [18].

The Rectified Linear Unit (ReLU) function is adopted as the activation function of hidden layers in both target and prediction networks. In contrast, a regression layer is added to the output layer. In order to train the network, the loss function is minimized using the widely-used Adam [75] algorithm.

Algorithm 1 displays the method by which GenCo-i evaluates bids before submitting. At first, GenCo-i determines whether to offer a random bid price with the probability of ϵ_i^t or to otherwise exploit collected knowledge and submit the best-known bid price. ϵ_i^t is computed similar to the Q-learning with decay algorithm.

Algorithm 1 Submitting a bid at time t by GenCo-i

1: $r \leftarrow \mathcal{U}(0, 1)$
2: **if** $r \leq \epsilon_i^t$ **then**
3: **if** $b_i^{t-k} = \ldots = b_i^{t-1}$ **then**
4: $b_i^t \leftarrow b_i^{t-1}$
5: **else**
6: $b_i^t \leftarrow$ choose a bid randomly from $\mathcal{B}_i$
7: **end if**
8: **else**
9: $b_i^* \leftarrow \underset{b_i^t \in B_i}{arg\,max}\{Q_i^t(s_i^t, b_i^t|\vec{w}_i) + \mu_i^t\}$
10: **end if**

To improve stability, lines 3-7 in Algorithm 1 do not allow GenCo-i to exercise its right for choosing a random bid if k previous bids are unchanged for some reason. The logic behind this strategy is that excessive exploration may induce disarray by undermining the coordination between agents when GenCos have already acquired enough knowledge [76].

When GenCo-i decides to submit the best-known bid, it feeds the current state, s_i^t, into the prediction network and chooses the bid that maximizes the reward according to the equation in line 9. In line 9, μ_i^t discourages altering GenCo's best-known bid if the Q-values of other options are just slightly better, i.e., $b_i^* \approx b_i^{t-1}$. The μ_i^t parameter also penalizes smaller bid prices than b_i^{t-1} to prevent a price war between GenCos.

To implement the proposed DQN model, the ConvNetSharp library [77] has been utilized in EMSimulator [78]. Moreover, EMSimulator employs the Microsoft Solver Foundation [79] library to clear the wholesale electricity market at each hour, through which the simulation process is accelerated by generating the DCOPF model on the fly. We assume that ISO uses a lookup table for the optimal solutions of previously solved problems. Doing so helps speed up the simulation even further at the expense of eliminating possible alternate optimal solutions.

3.4.3. Adjusting hyper-parameters

In this section, hyper-parameters are adjusted for both modified Q-learning and the proposed DQN model. A high-resolution simulation is conducted for 100,000 iterations to find the optimal ϵ_i^0 and α_i^0 values for the modified Q-learning algorithm. Each pair of settings is replicated 20 times to minimize the effect of random factors. A contour plot of the average total payoffs for the last 100 iterations is depicted in Figure 4 (a). As marked on the plot, the modified Q-learning algorithm performs considerably better when $\epsilon_i^0 = 0.9$ and $\alpha_i^0 = 0.1$. For the sake of fairness, we adopt these values for both DQN and the modified Q-learning since both algorithms follow the same recipe to decide when to explore new strategies or exploit collected information.

The developed DQN model has also multiple training-related hyper-parameters. Adam is usually considered as a fairly robust training algorithm to the choice of hyper-parameters; however, the learning rate may need to be adjusted according to the problem at hand [80]. As such, we run the DQN model for 20,000 iterations for various learning rates and plot the total loss function in Figure 4 (b). A learning rate of 0.01 provides a lower total loss at the end of the simulation. Accordingly, the following values are selected for hyper-parameters: a batch size of 32, a learning rate of 0.01, $\beta_1 = 0.99$, $\beta_2 = 0.999$, $\gamma = 0.7$, and weight decay of L_2 regularization of 0.015. The applied values for the exponential decay rates (i.e., β_1 and β_2) are shown empirically to be good choices in a wide variety of problems [81]. The γ and L_2 values are adopted from Ye et al. [24] to avoid huge weights.

4. Numerical experiments

4.1. Case study

Figure 5 illustrates a case study with seven nodes (i.e., regions) and four active agents, in which strong collusive strategies exist in 13 states. The presented case study is the extended version of the real Pennsylvania-New-Jersey-Maryland (PJM) five node power system, which is widely used in economic papers [82, 83, 22, 84] due to its simplicity. Finding a case study with analytical solutions was challenging, as we increased the complexity by having more nodes. Thus, we developed a script, which adjusts structure-related parameters to ensure the existence of SCEs according to Aliabadi et al. [20], given the set of bid prices. The availability of analytical solutions (e.g., Nash equilibria and SCEs) can assist us in examining the behavior of learning GenCos based on a game-theoretic framework. The complete list of all equilibria is presented in Appendix B.

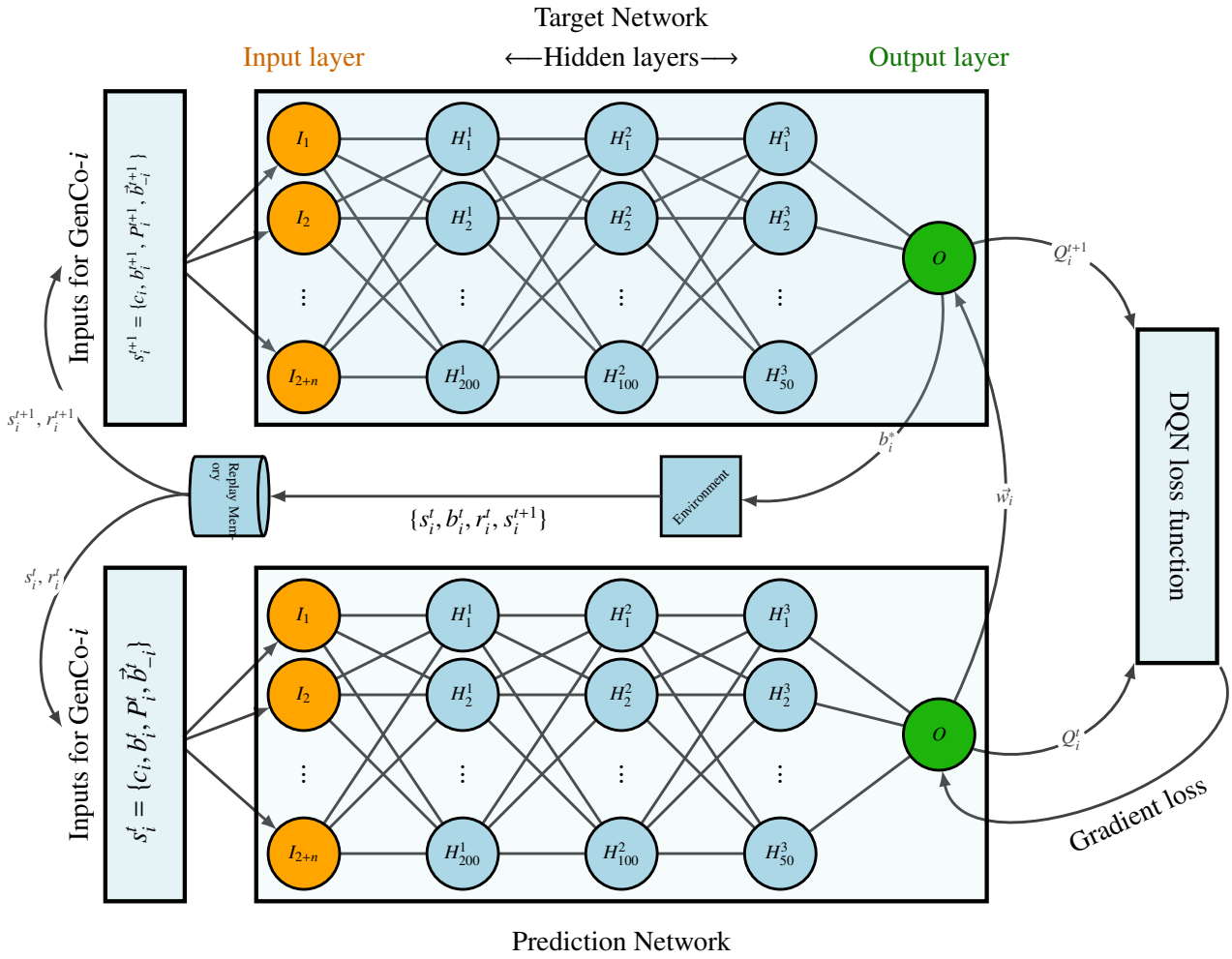
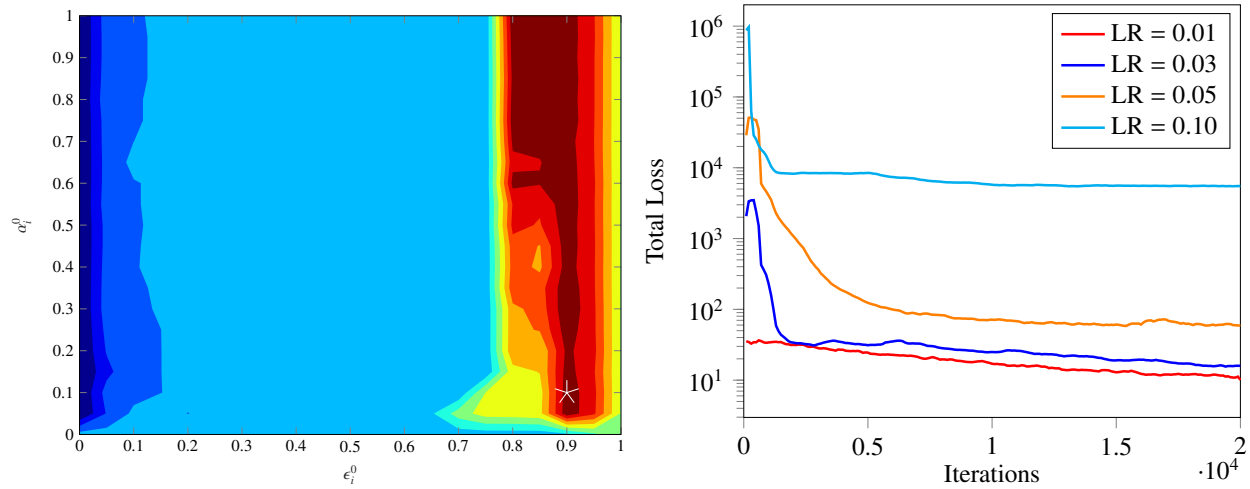


Figure 3: Schematic diagram of DQN model for GenCo-i. Input layers are painted in orange and output layers in green.



(a) α_i^0 and ϵ_i^0

(b) Adam's learning rate

Figure 4: Optimizing hyper-parameters for the modified Q-learning and the proposed DQN model. (a) the hotter the color, the higher the average of total payoffs. The selected setting $(\alpha_i^0, \epsilon_i^0)$ is marked with a star. (b) the total loss trends in logarithmic scale considering different learning rates.

The maximum generation capacity of GenCos (P_i^H) and load at demand centers (D_i) are written in MWh within the boundary of each node. Also, the maximum permissible flow between the source and destination nodes (F_{ij}) are mentioned next to the transmission lines. The dedicated set of bid prices for each GenCo, $\mathcal{B}_i$, and the unit cost of generating electricity, c_i are shown at the top-right corner. We devise bid prices such that no two GenCos have the same offer. Doing so will decrease the possibility of alternative optimal solutions per se.

4.2. Results

The simulation is conducted ten times over 100,000 iterations on a computer with 16 GB memory and an Intel Core i7-10510U processor. The program dedicates a thread with its exclusive memory space to each GenCo; hence, four logical cores out of eight are utilized thoroughly in this case study.

The initial recency (α_i^0) and initial exploration rates (ϵ_i^0) of all GenCos are set to 0.1 and 0.9, respectively. The prediction network is trained using the Adam algorithm as mentioned earlier. Figure 6 (a) shows the total loss ($\sum_i \mathcal{L}_i(\vec{w}_i)$) of the action-value function for 20,000 iterations and five replications. The pale lines represent different replications, and their average is drawn with a darker line.

Figure 6 (b) demonstrates payoff values of all GenCos as an instance when GenCos converge to an SCE (#12 in Table B.4). GenCo-2 and GenCo-6 gradually increase their payoffs while GenCo-1 and GenCo-5 struggle to hold their position in the market. These trends are due to the characteristics of payoffs under the Nash equilibria and SCEs. GenCo-6's payoffs are high under the Nash equilibria and SCEs, as well as GenCo-2 under SCEs; hence, when GenCo-2 and GenCo-6 exploit the best-known strategy more often, they can earn higher profits. GenCo-1 and GenCo-5, by contrast, have lower incomes under the Nash equilibria and most SCEs, which mandate bidding strategically to achieve the best possible outcome. Finally, all parties settle on an SCE strategy, ($b_1^{SCE} = 51, b_2^{SCE} = 47, b_5^{SCE} = 43, b_6^{SCE} = 34$), at around 87K, facilitated by Algorithm 1.

As depicted in Table 2, the converged tuple of bids using the proposed DQN outperforms the Nash equilibria and Q-learning with decay, in terms of payoffs. In fact, the t-test affirms that the two methods result in different total payoffs with the p-value of 0.002 considering a 95% confidence interval.

The bold rows mean convergence to an SCE occurs, as defined in Aliabadi et al. [20]. However, GenCo-i's average payoff ($\mathbb{E}[r_i^{DQN}]$) is not greater than the corresponding payoff in the SCE*. The proposed DQN algorithm was able to find an SCE in 70% of replications and the SCE* in 30% of occurrences. On average, DQN-equipped GenCos could earn €1466 per hour versus €1018 for Q-learning with decay. If all GenCos agree to act according to the SCE*, they could acquire €1654 per hour. This means that DQN-equipped GenCos could obtain 77.45% of the acquirable profit on average by diverging from the Nash equilibrium, with a total payoff of €820, to the SCE*.

As shown in Figure 7, all GenCos increased their average payoffs by using the DQN learning algorithm; however,

Table 2: Converged bid and payoff for each GenCo under two learning mechanisms

	DQN				Q-learning with decay			
#	b_1^*/r_1^*	b_2^*/r_2^*	b_5^*/r_5^*	b_6^*/r_6^*	b_1^*/r_1^*	b_2^*/r_2^*	b_5^*/r_5^*	b_6^*/r_6^*
1	**51/279**	**47/513**	**43/195**	**34/552**	46/182	32/324	33/96	34/0
2	**51/279**	**47/513**	**43/195**	**39/667**	41/147	32/324	33/27	29/437
3	51/217	42/594	53/207	49/897	41/430	37/328	48/0	34/552
4	**46/234**	**42/418**	**38/120**	**29/437**	36/328	32/270	38/0	29/437
5	**51/279**	**47/513**	**43/195**	**39/667**	36/387	37/0	33/160	34/43.2
6	**51/279**	**47/513**	**43/195**	**29/437**	41/430	37/328	43/0	34/552
7	**46/234**	**42/418**	**38/120**	**29/437**	36/112	32/324	33/96	34/552
8	**51/279**	**47/513**	**43/195**	**39/667**	36/112	32/324	33/96	34/552
9	31/367	32/116	33/0	29/437	36/112	32/324	33/96	34/552
10	51/217	42/594	43/117	34/552	41/147	32/324	43/117	34/552
$\mathbb{E}[r_i^*]$	266	471	154	575	239	287	69	423

GenCo-2 and GenCo-5 rise their payoffs significantly (i.e., $p-$ value < 0.05). The payoff of the fifth GenCo is particularly important since GenCo-5 is the player with minimum market power (see Table B.4). The red dashed lines represent the expected payoff when SCE* is played by all GenCos. As demonstrated, the number of hits on the red line (i.e., experiments with SCE* payoff) is increased meaningfully when DQN is exploited.

4.3. Sensitivity analysis on information availability

In the previous section, we have reported the results of the proposed DQN model when the submitted bid prices (b_i^t) are common knowledge; however, information disclosure varies extensively among countries. Therefore, in this section, we investigate GenCos' behavior when only λ_i^t is available immediately (i.e., DQN with LMP - DQNwL). To adopt this new change, we modify the state variable of the DQN model by replacing b_i^t with λ_i^t (i.e., $s_i^t = \{c_i, b_i^t, P_i^t, \vec{\lambda}_{-i}^t\}$).

The sample of outcomes exhibits that the probability of the game moving toward the Nash equilibria is 30% when GenCos are uncertain about rivals' bid prices; conversely, GenCos could not find SCEs in ten replications. Moreover, Figure 8 illustrates that the payoffs of GenCo-2, GenCo-5, and GenCo-6 are decreased significantly when the bid prices are hidden. The difference between payoffs can indicate the expected value of information for each GenCo.

The readers should note that the convergence to SCEs is not guaranteed since a simulation method is employed. All in all, we have a few salient observations:

1. The average payoff using DQN is higher than a time-dependent Q-learning method when bid prices are known.
2. Participants often receive payoffs larger than the Nash equilibria of the single-stage game, which might be caused by a price war and competition among players.
3. The proposed setting unveils the possibility of players unintentionally engaging in collusion in an oligopoly market.
4. The sensitivity analysis connotes the significance of knowledge about bid prices in establishing collusion.

5. Discussions

It is well-known in the literature that transparent markets facilitate maintaining tacit collusion via coordination of GenCos'

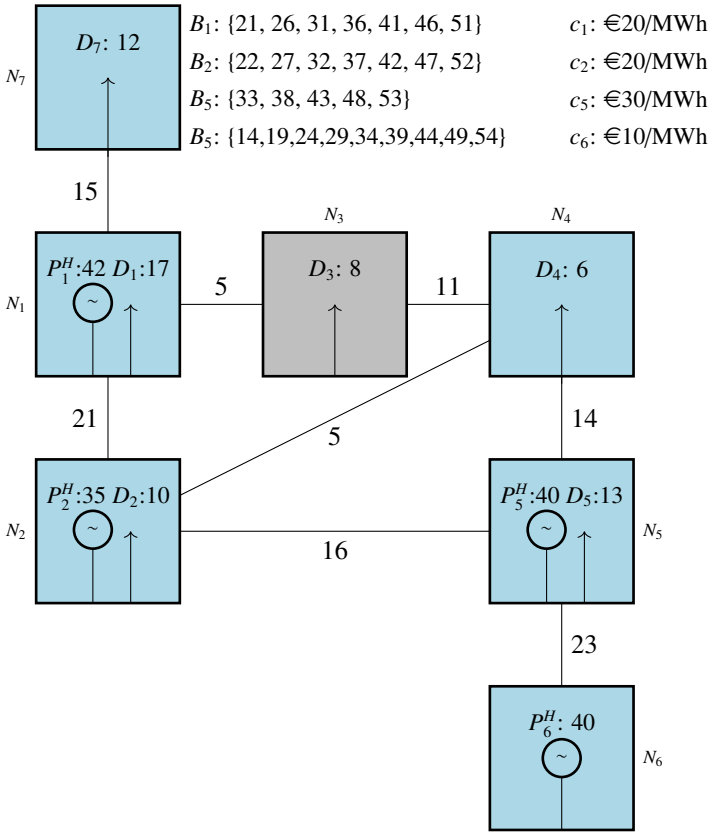
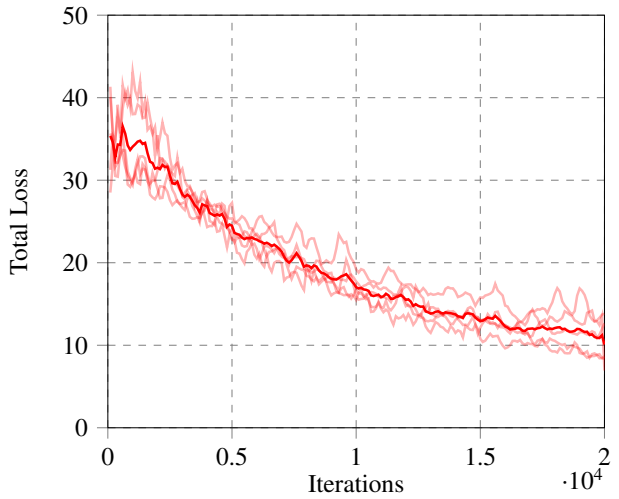
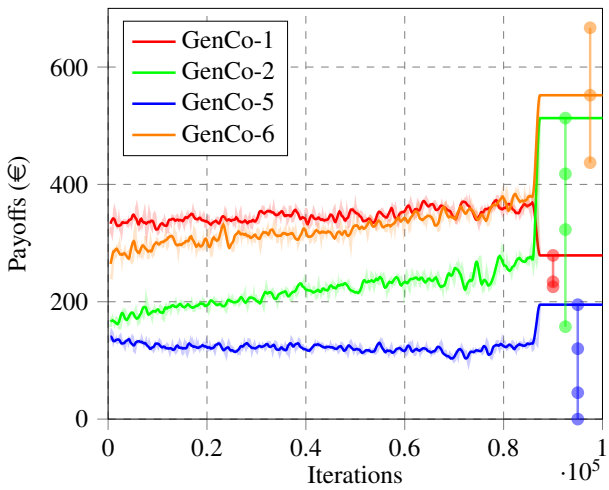


Figure 5: The case study with seven nodes and collusive strategies. The third node is considered as the base node.



(a) Loss reduction through learning



(b) GenCos' payoff transformation

Figure 6: (a) The light-toned lines represent various replications, and the dark line depicts their average. The total loss decreases as agents minimize the error using the Adam algorithm. (b) To clearly observe directions, payoffs are smoothed out in the depicted trends. Pale lines represent payoffs with a resolution of 100 iterations. The vertical lines, ranging from the Nash equilibria to SCE* for each GenCo, help the reader to put into perspective the converged payoff.

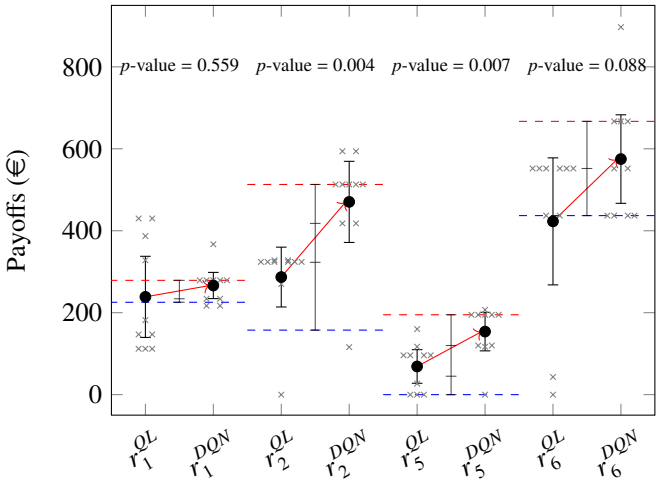


Figure 7: A 95% confidence interval for payoffs under the two learning algorithms. The blue and red dashed lines represent the payoff of the corresponding GenCo at the Nash equilibrium and SCE*, respectively. Other SCEs are displayed with black dash marks in between. A p-value < 0.05 means that the statistical difference between the average r_i^{QL} and the average r_i^{DQN} for GenCo-i is significant.

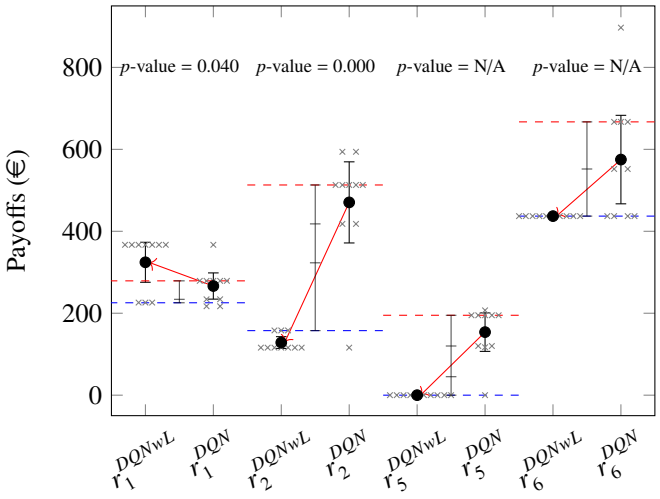


Figure 8: A 95% confidence interval for payoffs under the two information sets: DQN with LMPs (DQNwL) and DQN with bid prices. The blue and red dashed lines represent the payoff of the corresponding GenCo at the Nash equilibrium and SCE*, respectively. Other SCEs are displayed with black dash marks in between. A p-value < 0.05 means that the statistical difference between the average r_i^{DQNwL} and the average r_i^{DQN} for GenCo-i is significant. For GenCo-5 and GenCo-6, the p-values are unavailable due to zero standard deviation.

actions in repeated auctions [85, 86, 87, 88]. However, we designed a DQN model in this study, which has no information regarding the rivals' utilized technology, LMPs, or dispatched powers. The developed model discovers and sustains collusive strategies only by knowing the rivals' offered prices even though GenCos' objective is to improve the long-term payoff.

Although the proposed DQN model proves the possibility of tacit collusion among players in liberalized electricity markets, it should be noted that submitted bids are assumed as common knowledge. While GenCos may easily learn rivals' bid prices under pay-as-bid pricing, this information usually stays hidden behind the curtain of market-clearing prices under uniform and DCOPF pricing. Hence, for agents to collude using the proposed algorithm, the bidding curve should be (ideally) available immediately, which is not the case in many countries [89]. According to Wolak [90] and Yang et al. [91], information disclosure varies extensively among countries: some announce bidding curves instantly, while others release information with a delay of multiple weeks. When GenCos experience a disclosure delay of K hours, one possible approach is to train the proposed DQN model based on disclosed bid prices and assume the last announced hour as the current hour (see Figure 9). This approach can be reasonable when the network structure transforms sluggishly [92], and the influential players are often untouched in multiple years. As such, GenCos may still converge to a collusive strategy with a time delay. Another approach is to couple the proposed DQN model with other models, by which they can associate the bid prices to GenCos based on LMPs. Brown et al. [93] show that players can still identify the bid prices of a particular rival from the offered patterns even when the market de-identifies data.

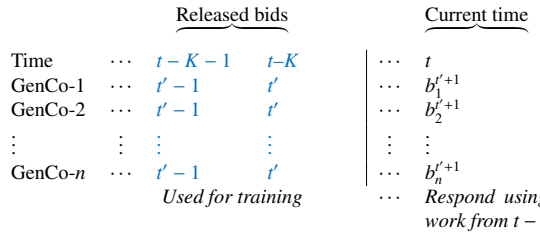


Figure 9: A possible training structure considering data disclosure delay of K hours.

There are also supporting arguments concerning the immediate release of bid prices by the ISO [94, 89, 95]. All in all, the general trend around the globe confirms that markets are moving toward full transparency, notably with data concerning historical bidding behaviors [28]. Hence, market designers and policymakers should consider the joint impact of autonomous pricing and information disclosure on GenCos' behavior prior to crafting market regulations.

To hinder GenCos from strategic bidding in liberalized markets, the ISO should diversify suppliers and distribute market power among smaller players [96]. For instance, the Turkish electricity market is becoming more diversified via the Renewable Energy Support Scheme (YEKDEM) and affordable photovoltaic cells [97, 98]. Governments can provide finan-

cial incentives for prosumers in local electricity markets to participate in the day-ahead markets and hedge against collusion [29]. Microgrids and local electricity markets are also suggested as efficient ways to integrate intermittent renewable energy sources with electricity markets [99]. However, market designers should constantly monitor small (local) markets to ensure healthy competition among players [32]. Finally, successful demand response strategies can relax the network congestion and provide a competitive environment, where GenCos would rather deviate from collusive strategies due to unattractive payoffs [100, 101, 102].

6. Conclusions and future work

Liberalized electricity markets should overcome empirical challenges to materialize predicted objectives completely. One major challenge is to achieve a fully competitive market by eliminating collusion of any type. Revealing the exercise of market power by participants is, in itself, a difficult task, and the use of autonomous pricing algorithms leads to added complexity. In this paper, we aim to investigate the impact of emerging DQN models on the behavior of players. The outcomes suggest that GenCos may be able to collude unintentionally while trying to ameliorate long-term profits. Therefore, policymakers and market designers should be vigilant regarding the combined effect of information disclosure and autonomous pricing, as new models exploit information more effectively.

Although the proposed DQN model does not need the solution to the DCOPF problem for rivals, it still requires knowing other GenCos' bidding curves. Consequently, one future research direction might be to employ off-the-shelf models (e.g., DDPG) or to design a new algorithm that only relies on publicly available information such as LMPs with monthly delay. Another research direction could be analyzing the effect of fluctuating demand, via a double-sided auction, on the learning capacity of GenCos when collusion among participants is possible.

Appendix A. DCOPF formulation

The DCOPF problem formulation [82, 20, 14] is given as follows:

$$\text{minimize}_{P_i^t,\theta_i^t} \qquad z^t = \sum_{i\in\mathcal{I}} b_i^t P_i^t \tag{A.1}$$

subject to

$$P_i^t - D_i^t = \sum_{ij\in\mathcal{A}} y_{ij}\left(\theta_i^t - \theta_j^t\right) \qquad \forall i \in \mathcal{I} \tag{A.2}$$

$$P_i^L \le P_i^t \le P_i^H \qquad \forall i \in \mathcal{I} \tag{A.3}$$

$$-\pi \le \theta_i^t \le \pi \qquad \forall i \in \mathcal{I} \tag{A.4}$$

$$|y_{ij}(\theta_i^t - \theta_j^t)| \le F_{ij}^H \qquad \forall ij \in \mathcal{A} \tag{A.5}$$

Here, D_i^t is the demand at node i and hour t, $\mathcal{A}$ is the set of available transmission lines, y_{ij} represents the admittance of the connecting line between a pair of nodes (i.e., i and j), P_i^L and P_i^H are the minimum and the maximum generation capacity of GenCo-i, respectively, and F_{ij}^H specifies the maximum permissible flow in the transmission line connecting node-i to node-j. For the sake of simplicity, we assume P_i^H, P_i^L, and D_i^t to be constant through time in the remainder of this paper.

The objective function in Eq. (A.1) is to minimize the electricity procurement cost. As stated earlier, P_i^t and θ_i^t are decision variables. Eq. (A.2) balances the flow of electricity by transmitting the extra power of each node into connected nodes. Eq. (A.3) confines the maximum and minimum permissible capacity of each GenCo. P_i^L can be set to a positive value when the power is already purchased or GenCo-i is selling according to a support mechanism such as feed-in tariffs. Eq. (A.4) limits the voltage angle within a finite range. Additionally, the value of θ_i^t at the reference node is set to zero. Finally, Eq. (A.5) controls the maximum flow through transmission lines. At the optimal solution, the dual variable corresponding to Eq. (A.2) sets the unit electricity price at each node (i.e., λ_i^t).

Appendix B. Equilibria of the Case Study

In this section, the analytical solution of the case study is presented based on Aliabadi et al. [20]. Table B.3 displays the tuple of bid prices (in €/MWh) and payoffs in Euro for the Nash equilibria. Table B.4 exhibits all SCEs with their corresponding payoffs. Among 2205 states, the most collusive state is at ($b_1^{SCE*} = 51, b_2^{SCE*} = 47, b_5^{SCE*} = 43, b_6^{SCE*} = 39$) with the payoff tuple of ($r_1^{SCE*} = 279, r_2^{SCE*} = 513, r_5^{SCE*} = 195, r_6^{SCE*} = 667$). The two Nash equilibria strategies are at ($b_1^N = 31, b_2^N = 27, b_5^N = \{33, 38\}, b_6^N = 29$) with the same payoff tuple ($r_1^N = 225.5, r_2^N = 157.5, r_5^N = 0, r_6^N = 437$). It is clear that $r_i^{SCE*} > r_i^N$, $\forall i \in \mathcal{I}$. As shown in the last column, the fifth GenCo always obtains the minimum payoff due to higher generation cost ($c_5 > c_1 = c_2 > c_6$).

Table B.3: Nash equilibria in the case study

#	$(b_1^N, b_2^N, b_5^N, b_6^N)$	$(r_1^N, r_2^N, r_5^N, r_6^N)$	Minimum Payoff
1	(31, 27, 33, 29)	(225.5, 157.5, 0, 437)	$r_5^N = 0$
2	(31, 27, 38, 29)	(225.5, 157.5, 0, 437)	$r_5^N = 0$

Under the Nash equilibria, transmission lines between nodes $2-4$ and $5-6$ become congested. Although GenCo-6 is capable enough to generate 40 MW per hour ($P_6^H = 40$ MWh), the dispatched power from GenCo-6 is capped to 23 MWh due to the transmission line constraint ($F_{5,6}^H = 23$). Congestion in the network causes the market cleared price of electricity to be node-dependent. For instance, under the Nash equilibria, the most expensive electricity is consumed at the fourth node ($\lambda_4^N =$ €39 per MWh) while the cheapest electricity is available at the second node ($\lambda_2^N =$ €27 per MWh).

Under the SCE* strategy, GenCos can increase the market cleared prices of all nodes ($\lambda_1^{SCE*} = 51 > \lambda_1^N = 31$, $\lambda_2^{SCE*} = 47 > \lambda_2^N = 27$, $\lambda_3^{SCE*} = 55 > \lambda_3^N = 35$, $\lambda_4^{SCE*} = 47 > \lambda_4^N = 39$, $\lambda_5^{SCE*} = 43 > \lambda_5^N = 33$, $\lambda_6^{SCE*} = 39 > \lambda_6^N = 29$, and $\lambda_7^{SCE*} = 51 > \lambda_7^N = 31$). Moreover, the transmission lines

between nodes $3-4$ and $4-5$ become congested under the SCE* strategy.

Table B.4: Calculated SCEs for the presented case study. The last row represents the most collusive state (SCE^*)

#	$(b_1^{SCE}, b_2^{SCE}, b_5^{SCE}, b_6^{SCE})$	$(r_1^{SCE}, r_2^{SCE}, r_5^{SCE}, r_6^{SCE})$	Minimum Payoff
1	(46, 37, 33, 29)	(234, 323, 45, 437)	$r_5^{SCE} = 45$
2	(46, 42, 33, 29)	(234, 418, 45, 437)	$r_5^{SCE} = 45$
3	(46, 42, 38, 29)	(234, 418, 120, 437)	$r_5^{SCE} = 120$
4	(46, 42, 38, 34)	(234, 418, 120, 552)	$r_5^{SCE} = 120$
5	(51, 42, 33, 29)	(279, 418, 45, 437)	$r_5^{SCE} = 45$
6	(51, 42, 38, 29)	(279, 418, 120, 437)	$r_5^{SCE} = 120$
7	(51, 42, 38, 34)	(279, 418, 120, 552)	$r_5^{SCE} = 120$
8	(51, 47, 33, 29)	(279, 513, 45, 437)	$r_5^{SCE} = 45$
9	(51, 47, 38, 29)	(279, 513, 120, 437)	$r_5^{SCE} = 120$
10	(51, 47, 38, 34)	(279, 513, 120, 552)	$r_5^{SCE} = 120$
11	(51, 47, 43, 29)	(279, 513, 195, 437)	$r_5^{SCE} = 195$
12	(51, 47, 43, 34)	(279, 513, 195, 552)	$r_5^{SCE} = 195$
13*	(51, 47, 43, 39)	(279, 513, 195, 667)	$r_5^{SCE} = 195$

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Availability of data and materials

All generated or analyzed data in this study are included in this manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author's contributions

DEA Conceptualization, Methodology, Visualization, Perform formal analysis, Investigation, Data curation, Coding, Writing - original draft, Writing - review & editing. **KC** Writing - Original draft.

Acknowledgments

The authors would like to thank the organizers of the 19th EPIA Conference on Artificial Intelligence, where the preliminary results of this manuscript were presented.

References

[1] H. Ritchie, M. Roser, Access to energy, Our World in Data. (2020). Retrieved Sep. 13, 2021 from `https://ourworldindata.org/energy-access`.

[2] U. Dubois, H. Meier, Energy affordability and energy inequality in Europe: Implications for policymaking, Energy Res Soc Sci 18 (2016) 21–35.

[3] P. L. Joskow, Introducing competition into regulated network industries: from hierarchies to markets in electricity, Ind. Corp. Change 5 (1996) 341–382.

[4] F. C. Özbuğday, B. Öğünlü, H. Alma, The sustainability of turkish electricity distributors and last-resort electricity suppliers: What did transition from vertically integrated public monopoly to regulated competition with privatized and unbundled firms bring about?, Util. Policy 39 (2016) 50–67.

[5] L. Papada, D. Kaliampakos, Measuring energy poverty in greece, Energy Policy 94 (2016) 157–165.

[6] J. J. Monast, Electricity competition and the public good: Rethinking markets and monopolies, U. Colo. L. Rev. 90 (2019) 667.

[7] A. K. David, F. Wen, Market power in electricity supply, IEEE Trans Energy Convers 16 (2001) 352–360.

[8] J. E. Harrington, Developing competition law for collusion by autonomous artificial agents, J Compet Law Econ 14 (2018) 331–363.

[9] S. Heim, G. Götz, Do pay-as-bid auctions favor collusion? evidence from Germany's market for reserve power, Energy Policy 155 (2021) 112308.

[10] BBC, 'big six' energy firms face competition inquiry, 2014. Retrieved Mar. 13, 2022 from `https://www.bbc.com/news/business-26734203`.

[11] A. Sweeting, Market power in the England and Wales wholesale electricity market 1995–2000, Econ J 117 (2007) 654–685.

[12] N. Fabra, J. Toro, Price wars and collusion in the Spanish electricity market, Int J Ind Organ 23 (2005) 155–181.

[13] X. Guan, Y.-C. Ho, D. L. Pepyne, Gaming and price spikes in electric power markets, IEEE Trans Power Syst 16 (2001) 402–408.

[14] E. Çelebi, G. Şahin, D. E. Aliabadi, Reformulations of a bilevel model for detection of tacit collusion in deregulated electricity markets, in: 2019 16th International Conference on the European Energy Market (EEM), IEEE, 2019, pp. 1–6.

[15] P. Razmi, M. O. Buygi, M. Esmalifalak, A machine learning approach for collusion detection in electricity markets based on Nash equilibrium theory, J Mod Power Syst Clean Energy (2020).

[16] E. Calvano, G. Calzolari, V. Denicolo, S. Pastorello, Artificial intelligence, algorithmic pricing, and collusion, Am Econ Rev 110 (2020) 3267–97.

[17] L. Bernhardt, R. Dewenter, Collusion by code or algorithmic collusion? when pricing algorithms take over, Eur Compet J 16 (2020) 312–342.

[18] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, Nature 518 (2015) 529–533.

[19] L. Chen, A. Mislove, C. Wilson, An empirical analysis of algorithmic pricing on Amazon marketplace, in: Proceedings of the 25th International Conference on World Wide Web, 2016, pp. 1339–1349.

[20] D. E. Aliabadi, M. Kaya, G. Şahin, Determining collusion opportunities in deregulated electricity markets, Electr Power Syst Res 141 (2016) 432–441.

[21] D. E. Aliabadi, M. Kaya, G. Şahin, An agent-based simulation of power generation company behavior in electricity markets under different market-clearing mechanisms, Energy Policy 100 (2017) 191–205.

[22] D. E. Aliabadi, M. Kaya, G. Sahin, Competition, risk and learning in electricity markets: An agent-based simulation study, Appl Energy 195 (2017) 1000–1011.

[23] M. B. Naghibi-Sistani, M. Akbarzadeh-Tootoonchi, M. J.-D. Bayaz, H. Rajabi-Mashhadi, Application of Q-learning with temperature variation for bidding strategies in market based power systems, Energy Conv Manag 47 (2006) 1529–1538.

[24] Y. Ye, D. Qiu, M. Sun, D. Papadaskalopoulos, G. Strbac, Deep reinforcement learning for strategic bidding in electricity markets, IEEE Trans Smart Grid 11 (2019) 1343–1355.

[25] A. Ebadi Torkayesh, Reformulations of a bi-level optimization problem detecting collusions in deregulated electricity markets, Master's thesis, Sabanci University, 2020.

[26] H. Kuhn, A. Tucker, Nonlinear programming, in: Proceedings of Second Berkeley Symposium on Mathematical Statistics and Probability, University of California Press, 1951, pp. 481–492.

[27] Y. Ye, D. Qiu, D. Papadaskalopoulos, G. Strbac, A deep Q network approach for optimizing offering strategies in electricity markets, in: 2019 International Conference on Smart Energy Systems and Technologies (SEST), IEEE, 2019, pp. 1–6.

[28] H. Guo, Q. Chen, Q. Xia, C. Kang, Deep inverse reinforcement learning for reward function identification in bidding models, IEEE Trans Power Syst (2021).

[29] D. E. Aliabadi, E. Çelebi, M. Elhüseyni, G. Şahin, Modeling, simulation, and decision support, in: Local Electricity Markets, Elsevier, 2021, pp. 177–197. doi:10.1016/B978-0-12-820074-2.00017-4.

[30] A. J. M. Kell, S. McGough, M. Forshaw, Machine learning applications for electricity market agent-based models: A systematic literature review, 2022. doi:10.48550/ARXIV.2206.02196.

[31] J. R. Vázquez-Canteli, Z. Nagy, Reinforcement learning for demand response: A review of algorithms and modeling techniques, Appl Energy 235 (2019) 1072–1089.

[32] P. Staudt, J. Gärttner, C. Weinhardt, Assessment of market power in local electricity markets with regards to competition and tacit collusion, in: Tagungsband Multikonferenz Wirtschaftsinformatik 2018, Leuphana University, 2018, pp. 912–923.

[33] I. T. Emami, H. A. Abyaneh, H. Zareipour, A. Bakhshai, A novel simulation-based method for assessment of collusion potential in wholesale electricity markets, Sustain Energy, Grids Netw 24 (2020) 100405.

[34] K. Poplavskaya, J. Lago, L. De Vries, Effect of market design on strategic bidding behavior: Model-based analysis of European electricity balancing markets, Appl Energy 270 (2020) 115130.

[35] K. Poplavskaya, J. Lago, S. Strömer, L. de Vries, Making the most of short-term flexibility in the balancing market: Opportunities and challenges of voluntary bids in the new balancing market design, Energy Policy 158 (2021) 112522.

[36] E. Namalomba, H. Feihu, H. Shi, Agent based simulation of centralized electricity transaction market using bi-level and Q-learning algorithm approach, Int J Electr Power Energy Syst 134 (2022) 107415.

[37] Q. Jia, Y. Li, Z. Yan, C. Xu, S. Chen, A reinforcement-learning-based bidding strategy for power suppliers with limited information, J Mod Power Syst Clean Energy (2021).

[38] T. Klein, Autonomous algorithmic collusion: Q-learning under sequential pricing, Rand J Econ 52 (2021) 538–558.

[39] A. G. Barto, S. Mahadevan, Recent advances in hierarchical reinforcement learning, Discret Event Dyn Syst-Theory Appl 13 (2003) 41–77.

[40] I. Erev, A. E. Roth, Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria, Am Econ Rev (1998) 848–881.

[41] M. Hemmati, M. Nili, N. Sadati, Reinforcement learning of heterogeneous private agents in a macroeconomic policy game, in: Progress in Artificial Economics, Springer, 2010, pp. 215–226.

[42] B. Avşar, D. E. Aliabadi, Parallelized neural network system for solving Euclidean traveling salesman problem, Appl Soft Comput 34 (2015) 862–873.

[43] T. Pinto, F. Falcão-Reis, Strategic participation in competitive electricity markets: Internal versus sectorial data analysis, Int J Electr Power Energy Syst 108 (2019) 432–444.

[44] D. E. Aliabadi, B. Avşar, R. Yousefnezhad, E. E. Aliabadi, Investigating global language networks using Google search queries, Expert Syst Appl 121 (2019) 66–77.

[45] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, et al., Mastering the game of Go with deep neural networks and tree search, Nature 529 (2016) 484–489.

[46] Y. Liang, C. Guo, Z. Ding, H. Hua, Agent-based modeling in electricity market using deep deterministic policy gradient algorithm, IEEE Trans. Power Syst. 35 (2020) 4180–4192.

[47] X. Pan, T. Zhao, M. Chen, S. Zhang, DeepOPF: A deep neural network approach for security-constrained dc optimal power flow, IEEE Trans. Power Syst. 36 (2020) 1725–1735.

[48] A. Velloso, P. Van Hentenryck, Combining deep learning and optimization for preventive security-constrained DC optimal power flow, IEEE Trans. Power Syst. 36 (2021) 3618–3628.

[49] M. Lehna, B. Hoppmann, R. Heinrich, C. Scholz, A reinforcement learning approach for the continuous electricity market of Germany: Trading from the perspective of a wind park operator, arXiv preprint arXiv:2111.13609 (2021).

[50] Y. Scholz, B. Fuchs, F. Borggrefe, K.-K. Cao, M. Wetzel, K. von Krbek, F. Cebulla, H. C. Gils, F. Fiand, M. Bussieck, T. Koch, D. Rehfeldt, A. Gleixner, D. Khabi, T. Breuer, D. Rohe, H. Hobbie, D. Schönheit, H. Ü. Yilmaz, E. Panos, S. Jeddi, S. Buchholz, Speeding up energy system models - a best practice guide, 2020. URL: https://elib.dlr.de/135507/.

[51] L. Nolting, T. Spiegel, M. Reich, M. Adam, A. Praktiknjo, Can energy system modeling benefit from artificial neural networks? application of two-stage metamodels to reduce computation of security of supply assessments, Comput Ind Eng 142 (2020) 106334.

[52] O. O. Owolabi, T. L. Schafer, G. E. Smits, S. Sengupta, S. E. Ryan, L. Wang, D. S. Matteson, M. G. Sherman, D. A. Sunter, Role of variable renewable energy penetration on electricity price and its volatility across independent system operators in the United States, arXiv preprint arXiv:2112.11338 (2021).

[53] M. Löschenbrand, Modeling competition of virtual power plants via deep learning, Energy 214 (2021) 118870.

[54] B. Esen, Utilizing genetic algorithm to detect collusive opportunities in deregulated energy markets, Master's thesis, Sabanci University, 2019.

[55] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, arXiv preprint arXiv:1509.02971 (2015).

[56] J. Wang, Conjectural variation-based bidding strategies with Q-learning in electricity markets, in: 2009 42nd Hawaii International Conference on System Sciences, IEEE, 2009, pp. 1–10.

[57] P. D. Klemperer, M. A. Meyer, Supply function equilibria in oligopoly under uncertainty, Econometrica (1989) 1243–1277.

[58] C. Ruiz, S. J. Kazempour, A. J. Conejo, Equilibria in futures and spot electricity markets, Electr. Power Syst. Res. 84 (2012) 1–9.

[59] Y. Chen, B. F. Hobbs, S. Leyffer, T. S. Munson, Leader-follower equilibria for electric power and NO_x allowances markets, Comput. Manag. Sci. 3 (2006) 307–330.

[60] M. Ventosa, R. Denis, C. Redondo, Expansion planning in electricity markets. two different approaches, in: Proceedings of the 14th Power Systems Computation Conference (PSCC), Seville, volume 26, 2002.

[61] C. J. Watkins, P. Dayan, Q-learning, Mach Learn 8 (1992) 279–292.

[62] A. G. Bakirtzis, A. C. Tellidou, Agent-based simulation of power markets under uniform and pay-as-bid pricing rules using reinforcement learning, in: 2006 IEEE PES Power Systems Conference and Exposition, IEEE, 2006, pp. 1168–1173.

[63] G. A. Antonopoulos, S. Vitiello, G. Fulli, M. Masera, Nodal pricing in the European internal electricity market, volume 30155, Publications Office of the European Union Luxembourg, 2020.

[64] R. Huisman, C. Huurman, R. Mahieu, Hourly electricity prices in day-ahead markets, Energy Econ 29 (2007) 240–248.

[65] A. C. Tellidou, A. G. Bakirtzis, Agent-based analysis of capacity withholding and tacit collusion in electricity markets, IEEE Trans Power Syst 22 (2007) 1735–1742.

[66] J. Sun, L. Tesfatsion, DC optimal power flow formulation and solution using QuadProgJ, Technical Report, Iowa State University, 2010. URL: https://dr.lib.iastate.edu/handle/20.500.12876/22480.

[67] D. Sharma, N. K. Yadav, G. Bhargava, A. Bala, Comparative analysis of ACOPF and DCOPF based LMP simulation with distributed loss model, in: 2016 International Conference on Control, Computing, Communication and Materials (ICCCCM), IEEE, 2016, pp. 1–6.

[68] K. Derinkuyu, F. Tanrisever, N. Kurt, G. Ceyhan, Optimizing day-ahead electricity market prices: increasing the total surplus for energy exchange istanbul, M&SOM-Manuf. Serv. Oper. Manag. 22 (2020) 700–716.

[69] E. Commission, D.-G. for Energy, A. Moser, N. Bracht, A. Maaz, Simulating electricity market bidding and price caps in the European power markets : S18 report, Publications Office, 2019. doi:doi/10.2833/252345.

[70] J. Liu, T. Lie, K. Lo, An empirical method of dynamic oligopoly behav-

ior analysis in electricity markets, IEEE Trans. Power Syst. 21 (2006) 499–506.
[71] R. Gibbons, et al., A primer in game theory (1992).
[72] UK Competition and Markets Authority, Pricing algorithms: Economic working paper on the use of algorithms to facilitate collusion and personalised pricing, 2018. Retrieved Sep. 13, 2021 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/746353/Algorithms_econ_report.pdf.
[73] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.
[74] R. Bellman, The theory of dynamic programming, Bull. Amer. Math. Soc. 60 (1954) 503–515.
[75] D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).
[76] W. Liu, Knowledge exploitation, knowledge exploration, and competency trap, Knowledge and Process Management 13 (2006) 144–161.
[77] A. Karpathy, Convnetsharp, 2016.
[78] D. Esmaeili Aliabadi, Analysis of collusion and competition in electricity markets using an agent-based approach, Ph.D. thesis, Sabanci University, 2016.
[79] Microsoft, Microsoft solver foundation, 2017. Available from https://www.nuget.org/packages/Microsoft.Solver.Foundation.
[80] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, MIT Press, 2016. http://www.deeplearningbook.org.
[81] J. Brownlee, Better deep learning: train faster, reduce overfitting, and make better predictions, Machine Learning Mastery, 2018.
[82] T. Krause, G. Andersson, D. Ernst, E. Vdovina-Beck, R. Cherkaoui, A. Germond, Nash equilibria and reinforcement learning for active decision maker modelling in power markets, in: Proceedings of the 6th IAEE European conference: modelling in energy economics and policy, 2004.
[83] T. Krause, G. Andersson, Evaluating congestion management schemes in liberalized electricity markets using an agent-based simulator, in: 2006 IEEE Power Engineering Society General Meeting, IEEE, 2006, pp. 8–pp.
[84] N. Mohammad, Y. Mishra, The role of demand response aggregators and the effect of gencos strategic bidding on the flexibility of demand, Energies 11 (2018) 3296.
[85] P. B. Overgaard, H. P. Møllgaard, Information exchange, market transparency and dynamic oligopoly, Technical Report 2007-3, University of Aarhus Economics, 2008.
[86] N.-H. M. von der Fehr, Transparency in electricity markets, Econ Energy Environ Policy 2 (2013) 87–106.
[87] P. Holmberg, F. A. Wolak, Electricity markets: Designing auctions where suppliers have uncertain costs, IFN Working Paper 1099, Research Institute of Industrial Economics (IFN), 2015. URL: http://hdl.handle.net/10419/129659.
[88] M.-C. Haufe, K.-M. Ehrhart, Auctions for renewable energy support–suitability, design, and first lessons learned, Energy Policy 121 (2018) 217–224.
[89] E. Lazarczyk, C. le Coq, Information disclosure in electricity markets, in: Heading Towards Sustainable Energy Systems: Evolution or Revolution?, 15th IAEE European Conference, Sept 3-6, 2017, International Association for Energy Economics, 2017.
[90] F. A. Wolak, Regulating competition in wholesale electricity supply, in: Economic regulation and Its reform, University of Chicago Press, 2014, pp. 195–290.
[91] Y. Yang, M. Bao, Y. Ding, Y. Song, Z. Lin, C. Shao, Review of information disclosure in different electricity markets, Energies 11 (2018).
[92] G. Gross, Transmission planning and investment in the competitive environment, in: 2005 IEEE Russia Power Tech, IEEE, 2005, pp. 1–3.
[93] D. P. Brown, A. Eckert, J. Lin, Information and transparency in wholesale electricity markets: evidence from Alberta, J Regul Econ 54 (2018) 292–330.
[94] A. Darudi, A. Z. Moghadam, H. J. D. Bayaz, Effects of bidding data disclosure on unilateral exercise of market power, in: 2015 International Congress on Technology, Communication and Knowledge (ICTCK), IEEE, 2015, pp. 17–24.
[95] J. Markard, E. Holt, Disclosure of electricity products-lessons from consumer research as guidance for energy policy, Energy Policy 31 (2003) 1459–1474.
[96] M. K. Heiman, B. D. Solomon, Power to the people: Electric utility restructuring and the commitment to renewable energy, Ann Assoc Am Geogr 94 (2004) 94–116.
[97] K. A. Çun, A review on development of renewable energy sector in Turkey in light of public policies and evaluation of YEKDEM mechanism effectiveness, Master's thesis, Sosyal Bilimler Enstitüsü, 2020.
[98] D. Esmaeili Aliabadi, D. Thrän, A. Bezama, B. Avşar, A systematic analysis of bioenergy potentials for fuels and electricity in Turkey: A bottom-up modeling, in: Transitioning to Affordable and Clean Energy, MDPI, 2022, pp. 295–234.
[99] N. Hatziargyriou, H. Asano, R. Iravani, C. Marnay, Microgrids, IEEE Power Energy Mag 5 (2007) 78–94.
[100] D. CooKE, Empowering customer choice in electricity markets, OECD, 2011. doi:10.1787/5kg3n27x4v41-en.
[101] R. Benjamin, Tacit collusion in electricity markets with uncertain demand, Rev. Ind. Organ. 48 (2016) 69–93.
[102] G. Yao, J. Wu, X. Liu, Gaming behavior in wholesale electricity markets with active demand response, in: 2017 13th IEEE Conference on Automation Science and Engineering (CASE), IEEE, 2017, pp. 1186–1190.