energies

MDPI

Article

Short-Term Electricity Futures Investment Strategies for Power Producers Based on Multi-Agent Deep Reinforcement Learning

Yizheng Wang [1,†], Enhao Shi [2,†], Yang Xu [3], Jiahua Hu [3] and Changsen Feng [2,*]

[1] Economic Research Institute of State Grid, Zhejiang Electric Power Company, Hangzhou 310000, China; sgcc_wangyz@163.com
[2] College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China; 202105031119@zjut.edu.cn
[3] State Grid Zhejiang Electric Power Co., Ltd., Hangzhou 310000, China; powermarket_zjsgcc@163.com (Y.X.); dreaminghu2009@163.com (J.H.)
* Correspondence: fcs@zjut.edu.cn
† These authors contributed equally to this work.

Abstract: The global development and enhancement of electricity financial markets aim to mitigate price risk in the electricity spot market. Power producers utilize financial derivatives for both hedging and speculation, necessitating careful selection of portfolio strategies. Current research on investment strategies for power financial derivatives primarily emphasizes risk management, resulting in a lack of a comprehensive investment framework. This study analyzes six short-term electricity futures contracts: base day, base week, base weekend, peak day, peak week, and peak weekend. A multi-agent deep reinforcement learning algorithm, Dual-Q MADDPG, is employed to learn from interactions with both the spot and futures market environments, considering the hedging and speculative behaviors of power producers. Upon completion of model training, the algorithm enables power producers to derive optimal portfolio strategies. Numerical experiments conducted in the Nordic electricity spot and futures markets indicate that the proposed Dual-Q MADDPG algorithm effectively reduces price risk in the spot market while generating substantial speculative returns. This study contributes to lowering barriers for power generators in the power finance market, thereby facilitating the widespread adoption of financial instruments, which enhances market liquidity and stability.

Keywords: electricity futures; price risk mitigation; power producer; multi-agent deep reinforcement learning; portfolio strategies

check for updates

Citation: Wang, Y.; Shi, E.; Xu, Y.; Hu, J.; Feng, C. Short-Term Electricity Futures Investment Strategies for Power Producers Based on Multi-Agent Deep Reinforcement Learning. *Energies* **2024**, *17*, 5350. https://doi.org/10.3390/en17215350

Academic Editor: William Holderbaum

Received: 30 September 2024
Revised: 16 October 2024
Accepted: 24 October 2024
Published: 28 October 2024

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

To manage price risk effectively, it is common to establish complementary markets for financial derivatives and introduce power financial derivatives. These strategies utilize hedging mechanisms to mitigate price risk. The development of power financial derivatives is a crucial element of the modern power financial market, driven by the high demand for risk hedging solutions [1,2]. The trading of power financial derivatives has become more standardized and transparent. Globally, countries have progressively established trading platforms for this market. For example, in Europe, Nordic power futures are primarily listed on the European Energy Exchange (EEX) and ICE ENDEX, while options contracts are transitioning from Nasdaq to EEX.

While participating in the spot power market, power producers can hedge risks, manage positions, and engage in arbitrage by using financial derivatives such as futures, options, and forward contracts within the power financial derivatives market [3]. The introduction of financial derivatives represents an inevitable trend in the evolution of the power market. Currently, a variety of financial derivatives for electricity, including futures, options, and swaps, are available globally, each designed to address different risk

management needs. Power producers in the power finance market must carefully select investment strategies to optimize risk hedging [4].

Current research on investment strategies for electricity financial derivatives primarily addresses uncertainty and risk factors. For example, Yucekaya employs Monte Carlo simulations to analyze scenarios involving spot, strike, and contract prices, with a focus on the Turkish electricity market. The study aims to optimize capacity allocation to maximize expected profit for a coal-fired unit within this market [5]. Sheybani et al. present an equilibrium model that incorporates market interactions, demand uncertainty, and consumer elasticity, examining the impact of strike and premium prices of put options on both the put option and day-ahead electricity markets [6]. Jaeck et al. explore the Samuelson effect across four major electricity futures markets (German, Nordic, Australian, and US) from 2008 to 2014, adding a new dimension to the effect [7]. Although these studies offer thorough analyses of price risk and investment strategies in the electricity financial market, they frequently lack specific scenarios for varying environments. Power financial transactions are inherently complex and variable, making it challenging to derive general scenarios from diverse trading conditions.

The emergence of artificial intelligence algorithms, particularly machine learning, has resulted in a significant increase in related methods within the power and finance markets. For instance, Zhipeng Liang et al. examine three advanced continuous reinforcement learning algorithms and introduce an adversarial training method that markedly enhances training efficiency, average daily return, and Sharpe ratio in backtests [8]. Shangkun Deng et al. propose a combined approach integrating XGBoost, SMOTE, and NSGA-II to forecast the directional movement of futures prices, demonstrating improved prediction accuracy, profitability, and return–risk ratio in simulation trading [9]. Germán G. Creamer et al. propose a multivariate distance nonlinear causality test (MDNC) that utilizes partial distance correlation within a time series framework, effectively detecting nonlinear lagged relationships and enhancing forecasting power when integrated with machine learning methods. This approach demonstrates a substantial improvement in forecasting key energy financial time series compared to the classical Granger causality method [10]. Additionally, Foued Saâdaoui et al. propose a novel multiscale machine learning forecasting model that employs Variational Mode Decomposition (VMD) to analyze cross-correlation between two European electricity markets, revealing significant long- and medium-term dependencies while indicating short-term independence. This model thus provides diverse investment opportunities [11].

Furthermore, numerous researchers have employed deep reinforcement learning (DRL) algorithms to devise timing strategies, stock-picking strategies, statistical arbitrage strategies, and portfolio strategies [12,13]. DRL can enhance investment strategies and real-time decision-making by interacting with the market and learning from feedback to refine optimal portfolio strategies. For example, Uyen Pham et al. investigate the application of DRL to develop an automated hedging strategy for Vietnam's stock market, aiming to improve trading performance and manage risk under volatile conditions [14]. Zhipeng Liang et al. examine three advanced continuous reinforcement learning algorithms and introduce an Adversarial Training method that significantly enhances training efficiency, average daily return, and Sharpe ratio in backtests [15]. Additionally, Xiao-Yang Liu et al. introduce the FinRL library, designed to assist beginners in developing and testing DRL trading strategies [16]. This library features a modular structure, incorporating state-of-the-art DRL algorithms and standard evaluation baselines, which facilitates strategy development, debugging, and comparison.

However, the DRL methods proposed for conventional financial derivatives trading strategies are not specifically tailored for the electricity trading process. Electricity, as an intangible commodity, operates under distinct market mechanics compared to traditional derivatives. Consequently, contracts in the electricity financial market differ significantly from those used in conventional financial derivatives, both in terms of structure and settlement mechanisms. Therefore, these studies may not fully address the unique characteristics and requirements of the electricity financial market. Specifically, the DRL methods discussed above do not yet provide adequate support for power producers engaging in the electricity finance market.

Based on this analysis, the paper proposes a multi-agent deep reinforcement learning (MADRL) algorithm that integrates various machine learning techniques to identify the optimal portfolio investment strategy for power producers in the electricity financial market. This approach acknowledges that financial derivatives, depending on their time granularity, exhibit varying degrees of effectiveness in risk hedging. Trading annual and quarterly contracts involves decision variability among power producers, influenced by their long-term investment objectives and specific circumstances. Given the complexity and diversity of strategies associated with these longer-term contracts, the focus shifts to short-term power financial derivatives, specifically day futures, week futures, and weekend futures. A cooperative environment is established by combining different machine learning techniques to develop tailored investment strategies for these derivatives. The traditional multi-agent deep deterministic policy gradient (MADDPG) algorithm within MADRL is enhanced by employing a dual-Q network in place of the target network in the actor to address the issue of overestimated Q values for the agent's actions. Subsequently, the dual-Q MADDPG algorithm is applied to historical data from the Nordic power market to identify the optimal portfolio strategy.

The numerical experiment results indicate that the Dual-Q MADDPG algorithm employed in this paper effectively mitigates spot market price risks. This is achieved through the cooperation of multiple agents and a combination of investments in various short-term electricity futures contracts. Additionally, this method generates objective market speculation income for power producers.

The contributions of this paper can be summarized as follows:

1. The MADRL method provides a portfolio investment strategy for power producers. It delivers an optimal buying and selling plan for futures by learning from historical trading data.
2. The algorithm accounts for the speculative and hedging behavior of power producers. Alongside hedging price risks with futures contracts, producers can also engage in futures transactions for additional income.
3. The traditional MADDPG algorithm is enhanced by using a double-target critic neural network in the actor-critic structure. This improvement helps power producers achieve greater prfit in the futures market.

2. Futures Trading Mechanisms for Power Producers

This paper uses the example of the EEX electricity futures market in Europe, which is a leading platform in the European power market. It offers trading in euro-denominated, cash-settled futures contracts across 20 European power markets. Among EEX's largest markets, the German Power Future stands out as the most liquid power product and serves as the benchmark contract in European wholesale power trading [17].

2.1. Types of Power Futures Contracts

European power futures are classified into day, week, weekend, quarter, and annual futures, along with other time-based variations. Regarding delivery mode, power futures are categorized into peak power futures, which cover high-demand periods, and base power futures.

We use the example of peak peak futures in Germany for illustration. According to the contract description provided by Intercontinental Exchange (ICE), a peak day load futures contract is financially settled based on the hourly prices from 08:00 to 20:00 CET of the German day-ahead auction in the control area operated by Amprion GmbH. This settlement applies to each day of the contract period, excluding weekends and irrespective of public holidays. For base day futures, the contract size is 1 MW per day, multiplied by the number of hours in the day, which can be 23, 24, or 25 h depending on the time of year (summer or winter) [18]. Similarly, base week futures cover a delivery period of 7 days, or 168 h, while weekend base futures cover a delivery period of 48 h on the weekend.We can list the delivery times for the six types of short-term futures contracts as shown in Table 1.

Table 1. The contract size of the six types of short-term futures.

Contract Type	Delivery Period	Delivery Time
peak	day	12 h
	weekend	24 h
	week	84 h
base	day	24 h [1]
	weekend	48 h
	week	168 h

[1] We use a 24 h time format to avoid the impact of daylight saving time adjustments.

2.2. The Whole Trading Cycle

Futures trading cycles are typically divided into two phases: the contract period and the delivery period. During the contract period, traders establish positions in futures contracts within the trading hours set by EEX. In the delivery period, the holder of a long position in a futures contract is obligated to execute the contract. The detailed process is illustrated in Figure 1.

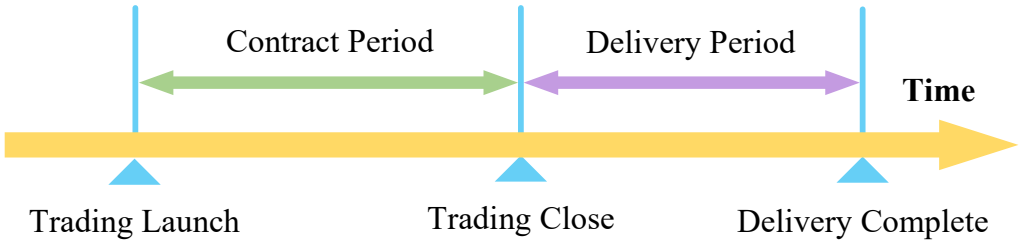


Figure 1. The futures contracts trading process.

Power producers can engage in futures contracts during the contract period to lock in the price of electricity delivery and pay a specified deposit. If the power producer does not sell the contract before the delivery date, it is obligated to fulfill the contract terms. If the agreed delivery price in the contract exceeds the spot market price, the power producer stands to make a profit; conversely, if it is lower, the producer will incur a loss. Additionally, power producers have the option to sell their contracts before the delivery date, potentially realizing a profit. This option entails a decision-making process for the power producers.

2.3. Decision-Making of Power Producers

Power producers engage in the power futures market for three primary reasons:

- Hedging: Power producers mitigate the impact of fluctuations in spot market electricity prices by using futures contracts. These contracts allow them to lock in electricity prices in advance, thereby reducing the risk of losses from excessively low prices [19]
- Speculation: Power producers can sell futures contracts before expiration without having to take delivery of the electricity. If the price of the futures contract has increased by that time, they will realize a profit.
- Arbitrage: Power producers can profit from price discrepancies across different financial markets using the same futures contract. However, this topic is beyond the scope of the current article [20].

Next, an illustrative example is provided. A power producer anticipates a significant drop in the market price of electricity during the third week of the month due to expected changes in energy supply and demand, among other factors. Consequently, the producer purchases a base week futures contract at a price of EUR 100/MWh. The producer's cost of generation is EUR 80/MWh, while the average spot market price during the delivery period is EUR 60/MWh. Despite the spot market price being lower than the generation cost, the producer still realizes a profit due to the favorable futures contract price. The net profit earned by the power producer during the 168 h delivery period of the base week futures contract is

$$Pure\ Profit = (100 - 80) \times 168 \times 50 = 168,000\ EUR. \quad (1)$$

Meanwhile, the power producer avoided the following losses:

$$Avoid\ Loss = (80 - 60) \times 168 \times 50 = 168,000\ EUR. \quad (2)$$

Overall, power producers achieve a profit of EUR 336,000 by participating in the futures market.

Similarly, if the power producer opts to purchase 50 peak week futures contracts at a price of EUR 120/MWh, and the average electricity price during the peak delivery period is EUR 70/MWh, the profit earned by the power producer is

$$profit = (120 - 70) \times 84 \times 50 = 210,000\ EUR. \quad (3)$$

In addition to hedging their risks, power producers can also engage in speculation within the futures market. Since futures contracts can be sold short at any time before delivery, power producers have the opportunity to realize speculative profits. For example, a power producer can achieve a profit by purchasing a base day futures contract at EUR 80/MWh and subsequently selling 100 contracts at EUR 90/MWh before the delivery period.

$$profit = (90 - 80) \times 24 \times 100 = 24,000\ EUR \quad (4)$$

However, if the power producer sells the futures contract, this action may impact the effectiveness of subsequent risk hedging.

3. Materials and Methods

Based on the analysis in the previous section, power producers typically face two primary challenges when participating in the electricity futures market:

1. Determine whether to purchase new contracts or sell existing ones, and establish the quantity to be bought or sold.
2. Formulate a strategic investment approach for various futures contracts to maximize profits.

To address these two challenges, a MADRL approach can be employed. This method leverages multiple agents to develop distinct investment strategies for futures contracts, which are subsequently integrated into a cohesive investment strategy.

3.1. Overview of MADRL

General planning theory typically addresses discrete subproblems, while Reinforcement Learning (RL) focuses on the overarching challenge of an agent interacting with an uncertain environment [21]. In RL, agents receive rewards based on specific objectives by perceiving the environment and selecting actions that influence it. Deep Learning (DL) has significantly advanced RL by enhancing state–action mappings, thereby improving the agent's capacity to comprehend the environment and make effective decisions. The integration of DL with RL has led to the emergence of DRL, which combines the strengths of both approaches [22].

In real-world applications, multiple agents often learn and manage various tasks simultaneously, resulting in the emergence of the MADRL problem. However, increasing the number of agents presents challenges in managing their interactions. In MADRL, even without direct communication, the actions of one agent can still impact others. For example, the decisions made by one trader can influence stock prices, potentially benefiting or adversely affecting other automated traders.

In MADRL, there are four primary types of relationships between agents: Fully Cooperative, Fully Competitive, Mixed Cooperative and Competitive, and Self-Interested. When multiple agents engage in a cooperative relationship, the problem can often be modeled as a Decentralized Partially Observable Markov Decision Process (Dec-POMDP), which can be formulated as a tuple $\langle \mathcal{I}, \mathcal{S}, \mathcal{A}, \mathcal{O}, r, \gamma \rangle$, where:

- $\mathcal{I}$ denotes the set of agents;
- $\mathcal{S}$ denotes the set of observations;
- $\mathcal{A}$ represents the joint action set;
- $\mathcal{O}$ represents the joint observation set;
- r is the reward function applicable to all agents;
- γ is the discount factor.

This paper models the investment strategies of power producers through six distinct agents: Agent 1 focuses on base day futures; Agent 2 concentrates on base week futures; Agent 3 specializes in base weekend futures; Agent 4 addresses peak day futures; Agent 5 focuses on peak week futures; and Agent 6 specializes in peak weekend futures.

The MADRL framework employs Centralized Training with Distributed Execution (CTDE). As illustrated in Figure 2, each agent corresponds to different specifications of futures contracts. During offline centralized training, the agents share observational information, while during execution, each agent independently determines the appropriate trading volume for the futures contracts.

3.2. Action Sets and Action Spaces

Futures contracts are typically listed several days or weeks before the delivery date, and different types of contracts have varying delivery periods. Therefore, we can assume the following:

- l : Day futures contracts will be available l days before the delivery date;
- m : week futures contracts will be available l weeks before the delivery date;
- n : weekend futures contracts will be available n weekends before the delivery date.

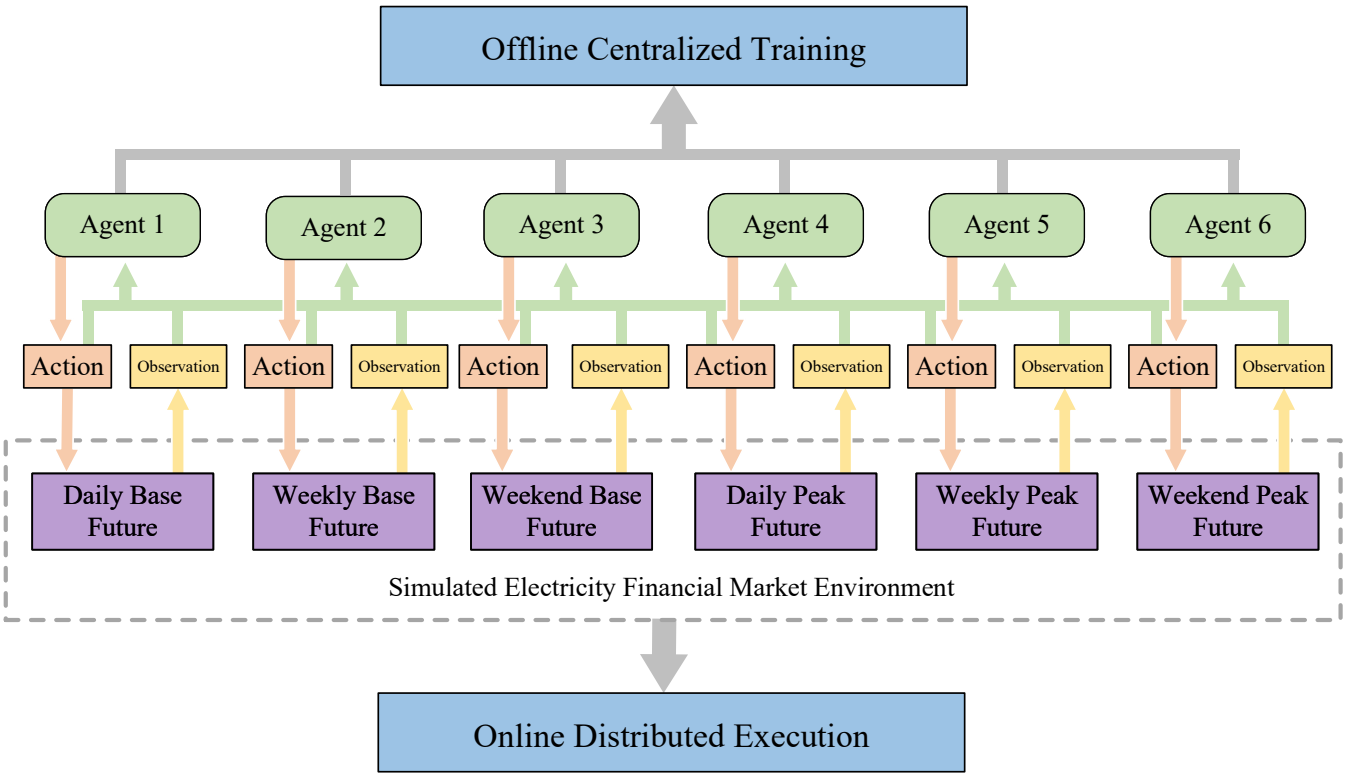


Figure 2. CTDE framework of six agents.

Thus, each trading day can determine the corresponding contract for l days, m weeks, and n weeks back. The action sets available to the six agents are as follows:

$$A_1 = \left\{ v_{i,t}^{b,d} \right\} \quad i = 1, 2, ..., l, \tag{5}$$

$$A_2 = \left\{ v_{i,t}^{b,w} \right\} \quad i = 1, 2, ..., m, \tag{6}$$

$$A_3 = \left\{ v_{i,t}^{b,k} \right\} \quad i = 0, 2, ..., n, \tag{7}$$

$$A_4 = \left\{ v_{i,t}^{p,d} \right\} \quad i = 1, 2, ..., l, \tag{8}$$

$$A_5 = \left\{ v_{i,t}^{p,w} \right\} \quad i = 1, 2, ..., m, \tag{9}$$

$$A_6 = \left\{ v_{i,t}^{p,k} \right\} \quad i = 0, 2, ..., n, \tag{10}$$

where $v_{i,t}^{b,d}$ is the volume of base day futures traded after i day. $v_{i,t}^{b,w}$ is the volume of base week futures traded after i week. $v_{i,t}^{b,k}$ represents the volume of base weekend futures traded after i weekends, where $i = 0$ denotes the weekend of the current week. Specifically, futures contracts delivered during this weekend are tradable from Monday through Friday of the same week. $v_{i,t}^{p,d}$, $v_{i,t}^{p,w}$, and $v_{i,t}^{p,k}$ denote the trading volumes of peak day, week, and weekend futures contracts, respectively.

At the time t, Agent 1 must determine the purchase or sale volumes of l types of base day futures. Similarly, Agent 2 must decide the purchase or sale volumes of m types of week futures, and Agent 3 must determine the purchase or sale volumes of n types of weekend futures. Agents 4, 5, and 6 perform analogous actions concerning peak futures. The meanings of the symbols $v_{1,t}^{b,d}$, $v_{1,t}^{p,d}$, $v_{1,t}^{b,w}$, $v_{1,t}^{p,w}$, $v_{1,t}^{b,k}$, and $v_{1,t}^{p,k}$ in the action set are illustrated in Figure 3.

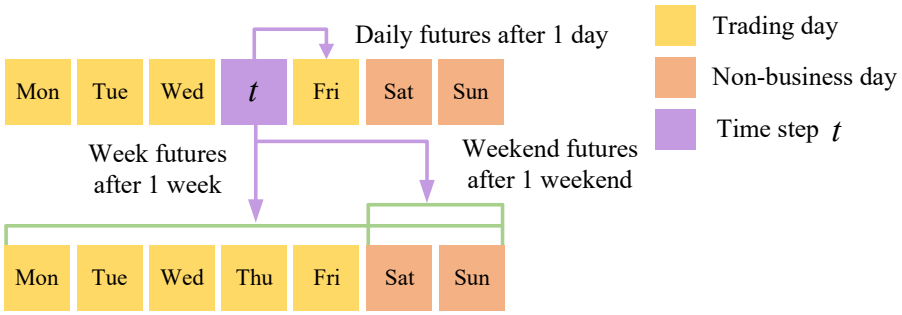


Figure 3. The action set of agents.

Next, we discuss the action space of the agents. Since the signing of futures contracts requires the payment of a corresponding margin, there is a limit on the number of contracts that can be signed and traded daily. The upper limit of an agent's actions depends on the power producer's capital position. Therefore, we further define the total cost for the power producer to trade futures contracts as

$$\begin{aligned} C_t = &\sum_{i=1}^{l}\left(m_{i,t}^{b,d}v_{i,t}^{b,d} + m_{i,t}^{p,d}v_{i,t}^{p,d}\right) \\ &+ \sum_{i=1}^{m}\left(m_{i,t}^{b,w}v_{i,t}^{b,w} + m_{i,t}^{p,w}v_{i,t}^{p,w}\right) + \sum_{i=1}^{n}\left(m_{i,t}^{b,k}v_{i,t}^{b,k} + m_{i,t}^{p,k}v_{i,t}^{p,k}\right), \end{aligned} \tag{11}$$

where $m_{i,t}^{b,d}$ and $m_{i,t}^{p,d}$ are the margins for base and peak day futures, respectively. $m_{i,t}^{b,w}$ and $m_{i,t}^{p,w}$ are the margins for base and peak week futures, respectively. $m_{i,t}^{b,k}$ and $m_{i,t}^{p,k}$ are the margins for base and peak weekend futures, respectively.

Power producers must allocate less than their maximum available funds when purchasing futures contracts.

$$C_t \leq M_{t,\max} \tag{12}$$

where $M_{t,\max}$ represents the maximum amount of money a power producer can invest in futures at time t.

As mentioned in Section 2.3, power producers can sell futures contracts for speculative profit before the delivery date. When the trading volume of futures $v < 0$, it indicates that power producers are partially selling their futures holdings. Therefore, the lower bound of an agent's action at time τ can be defined as

$$v_{i,\tau}^{b,d} \geq -\sum_{t=1}^{\tau-1} v_{i,t}^{b,d}, \tag{13}$$

$$v_{i,\tau}^{b,w} \geq -\sum_{t=1}^{\tau-1} v_{i,t}^{b,w}, \tag{14}$$

$$v_{i,\tau}^{b,k} \geq -\sum_{t=1}^{\tau-1} v_{i,t}^{b,k}, \tag{15}$$

$$v_{i,\tau}^{p,d} \geq -\sum_{t=1}^{\tau-1} v_{i,t}^{p,d}, \tag{16}$$

$$v_{i,\tau}^{p,w} \geq -\sum_{t=1}^{\tau-1} v_{i,t}^{p,w}, \tag{17}$$

$$v_{i,\tau}^{p,k} \geq -\sum_{t=1}^{\tau-1} v_{i,t}^{p,k}, \tag{18}$$

where we assume that when a futures contract is executed, the trading volume of futures $v = 0$.

Since power producers often hold multiple futures contracts at varying prices, they must determine which contracts to sell and the quantities to sell when engaging in speculative trading for profit. This requires making assumptions about the rules governing the sale of futures and options contracts by power producers. It is assumed that when a power producer holds contracts at varying prices and decides to close a futures or options position, the contract with the largest spread will be sold preferentially to maximize short-term profits.

In addition, power producers must trade futures during the trading days specified by the EEX. Trades typically occur on business days that exclude weekends and holidays. For the purposes of this paper, we assume that trading takes place from Monday to Friday each week. Therefore, on Saturdays and Sundays, the trading volume of futures is $v = 0$.

3.3. Observation Sets and Observation Spaces

The observation sets available to the six agents can be described as follows:

$$O_1 = \left\{ \lambda_t^b, c_{i,t}^{b,d}, m_{i,t}^{b,d} \right\} \quad i = 1, 2, ..., l, \tag{19}$$

$$O_2 = \left\{ \lambda_t^b, c_{i,t}^{b,w}, m_{i,t}^{b,w} \right\} \quad i = 1, 2, ..., m, \tag{20}$$

$$O_3 = \left\{ \lambda_t^b, c_{i,t}^{b,k}, m_{i,t}^{b,k} \right\} \quad i = 1, 2, ..., n, \tag{21}$$

$$O_4 = \left\{ \lambda_t^p, c_{i,t}^{p,d}, m_{i,t}^{p,d} \right\} \quad i = 1, 2, ..., l, \tag{22}$$

$$O_5 = \left\{ \lambda_t^p, c_{i,t}^{p,w}, m_{i,t}^{p,w} \right\} \quad i = 1, 2, ..., m, \tag{23}$$

$$O_6 = \left\{ \lambda_t^p, c_{i,t}^{p,k}, m_{i,t}^{p,k} \right\} \quad i = 1, 2, ..., n, \tag{24}$$

where λ_t^b is the average electricity price in the base hours at the time t. λ_t^p is the average electricity price in the peak hours at the time t. $c_{i,t}^{b,d}$ and $c_{i,t}^{p,d}$ are the contract strike price for base and peak day futures, respectively. $c_{i,t}^{b,w}$ and $c_{i,t}^{p,w}$ are the contract strike price for base and peak week futures, respectively. $c_{i,t}^{b,k}$ and $c_{i,t}^{p,k}$ are the contract strike price for base and peak weekend futures, respectively. These observation spaces provide fundamental information about futures contracts and will enhance the agent's decision-making.

3.4. Rewards Functions

The MADRL algorithm discussed in this paper operates within a cooperative multi-agent setting, where all agents share a common reward signal that reflects their joint contributions to the system's performance. The reward for the agents is derived from the overall profits of the power producer, as these profits effectively capture the collective performance and operational efficiency of the agents. The profit gained by power producers consists of five components: the profit from speculating on futures contracts, the profit gained from the delivery of futures, the recovered contract margin, the margin paid for signing the futures contract, and the return of the margin gained at the last time step.

3.4.1. Profit from Speculating

A power producer may sell its holdings of futures contracts prior to the delivery period. At this point, the power producer is considered an ordinary participant in the futures market. Once the contract is sold, the power producer has no obligation to execute it. If the market price of the futures contract at that time is higher than the purchase price, the power producer can realize a profit. Additionally, by selling the futures contract, the previously paid margin can be recovered.

We take the example of base day futures. When $v_{i,t}^{b,d} \leq 0$, it means that the contract is sold at the current time. The power producer's portion of the gain can be expressed as follows:

$$r_{spec}^{b,d} = \sum_{j=1}^{i} \sum_{\tau=0}^{t} \left[24\left(c_{j,\tau}^{b,d} - c_{i,t}^{b,d} \right) + m_{j,\tau}^{b,d} \right] \left| v_{j,\tau}^{b,d} \right| \quad \forall j + \tau = i + t, \tau \leq t, \tag{25}$$

$$\sum_{j=1}^{i} \sum_{\tau=0}^{t} v_{j,\tau}^{b,d} = v_{i,t}^{b,d} \quad \forall j + \tau = i + t, \tau \leq t, \tag{26}$$

where $r_{spec}^{b,d}$ denotes the power producer's speculative profit on base day futures. $c_{j,\tau}^{b,d}$ denotes the price of a futures contract purchased on day τ for delivery after day j. $v_{j,\tau}^{b,d}$ denotes the volume. When $j + \tau = i + t$ and $\tau \leq t$, it means it is a contract for delivery on the same day. The number 24 in Equation (25) refers to the 24 h delivery period of the base load futures contract on that day. Equation (26) refers to the amount of futures contracts sold for all held futures contracts, equal to the action of the current smart body.

Similarly, the speculative returns on the peak day futures contracts can be expressed as follows:

$$r_{spec}^{p,d} = \sum_{j=1}^{i} \sum_{\tau=0}^{t} \left[12\left(c_{j,\tau}^{p,d} - c_{i,t}^{p,d} \right) + m_{j,\tau}^{p,d} \right] \left| v_{j,\tau}^{p,d} \right| \quad \forall j + \tau = i + t, \tau \leq t, \tag{27}$$

$$\sum_{j=1}^{i} \sum_{\tau=0}^{t} v_{j,\tau}^{p,d} = v_{i,t}^{p,d} \quad \forall j + \tau = i + t, \tau \leq t, \tag{28}$$

where $r_{spec}^{p,d}$ denotes the power producer's speculative profit on peak day futures.

When it comes to weeks, we need to use a downward rounding function $\lfloor t/7 \rfloor$ to calculate the week to which time t belongs. When $j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor$, it means that the two-week futures contracts are delivered in the same week. Therefore, the speculative returns on the remaining four futures contracts can be expressed as follows:

$$r_{spec}^{b,w} = \sum_{j=1}^{i} \sum_{\tau=1}^{t} \left[24\left(c_{j,\tau}^{b,w} - c_{i,t}^{b,w} \right) + m_{j,\tau}^{b,w} \right] \left| v_{j,\tau}^{b,w} \right| \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{29}$$

$$r_{spec}^{p,w} = \sum_{j=1}^{i} \sum_{\tau=1}^{t} \left[12\left(c_{j,\tau}^{p,w} - c_{i,t}^{p,w} \right) + m_{j,\tau}^{p,w} \right] \left| v_{j,\tau}^{p,w} \right| \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{30}$$

$$r_{spec}^{b,k} = \sum_{j=1}^{i} \sum_{\tau=1}^{t} \left[24\left(c_{j,\tau}^{b,k} - c_{i,t}^{b,k} \right) + m_{j,\tau}^{b,k} \right] \left| v_{j,\tau}^{b,k} \right| \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{31}$$

$$r_{spec}^{p,k} = \sum_{j=1}^{i} \sum_{\tau=1}^{t} \left[12\left(c_{j,\tau}^{p,k} - c_{i,t}^{p,k} \right) + m_{j,\tau}^{p,k} \right] \left| v_{j,\tau}^{p,k} \right| \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{32}$$

$$\sum_{j=1}^{i} \sum_{\tau=1}^{t} v_{j,\tau}^{b,w} = v_{i,t}^{b,w} \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{33}$$

$$\sum_{j=1}^{i} \sum_{\tau=1}^{t} v_{j,\tau}^{p,w} = v_{i,t}^{p,w} \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{34}$$

$$\sum_{j=1}^{i} \sum_{\tau=1}^{t} v_{j,\tau}^{b,k} = v_{i,t}^{b,k} \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{35}$$

$$\sum_{j=1}^{i} \sum_{\tau=1}^{t} v_{j,\tau}^{p,k} = v_{i,t}^{p,k} \quad \forall j + \lfloor \tau/7 \rfloor = i + \lfloor t/7 \rfloor, \tau \leq t, \tag{36}$$

where $r^{b,w}_{spec}$, $r^{p,w}_{spec}$, $r^{b,k}_{spec}$, and $r^{p,k}_{spec}$ denote the power producer's speculative profit on base week, peak week, weekend base, snf weekend peak futures respectively.

The total speculative profits from the six futures contracts represent the profit made by the power producers through speculation on futures:

$$r_{spec} = r^{b,d}_{spec} + r^{b,w}_{spec} + r^{b,k}_{spec} + r^{p,d}_{spec} + r^{p,w}_{spec} + r^{p,k}_{spec}. \tag{37}$$

3.4.2. Profit from the Delivery

When a contractual delivery is made, the power producer is obliged to sell electricity based on the contract price. According to the classification in Table 1, day futures can be divided into base futures and peak futures, so the rewards also include the returns from both types of futures. At the time t, The profits of a power producer engaged in day futures trading can be expressed as follows:

$$r^{b,d}_{deliv} = \sum_{i=1}^{l} 24\left(c^{b,d}_{i,t-i} - \lambda_{b,t}\right) v^{b,d}_{i,t-i}, \tag{38}$$

$$r^{p,d}_{deliv} = \sum_{i=1}^{l} 12\left(c^{p,d}_{i,t-i} - \lambda_{p,t}\right) v^{p,d}_{i,t-i}, \tag{39}$$

where $r^{b,d}_{deliv}$ and $r^{p,d}_{deliv}$ denote the power producer's delivery profit on base day and peak day futures respectively. When $c^{b,d}_{i,t-i} \leq \lambda_{b,t}$ or $c^{p,d}_{i,t-i} \leq \lambda_{p,t}$, the power producer received positive profit, which means that is successfully hedges its price risk through the futures contract.

Similarly, the profit on delivery of week and weekend futures can be expressed as follows:

$$p^{b,w}_{deliv} = \sum_{i=1}^{m} \sum_{j=1}^{7m+(t \bmod 7)} 24\left(c^{b,w}_{i,t-j} - \lambda_{b,t}\right) v^{b,w}_{i,t-j} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{40}$$

$$p^{p,w}_{deliv} = \sum_{i=1}^{m} \sum_{j=1}^{7m+(t \bmod 7)} 12\left(c^{p,w}_{i,t-j} - \lambda_{p,t}\right) v^{p,w}_{i,t-j} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{41}$$

$$p^{b,k}_{deliv} = \sum_{i=0}^{n} \sum_{j=0}^{7n+(t \bmod 7)} 24\left(c^{b,k}_{i,t-j} - \lambda_{b,t}\right) v^{b,k}_{i,t-j} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{42}$$

$$p^{p,k}_{deliv} = \sum_{i=0}^{n} \sum_{j=0}^{7n+(t \bmod 7)} 12\left(c^{p,k}_{i,t-j} - \lambda_{p,t}\right) v^{p,k}_{i,t-j} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{43}$$

where mod is the residual function. $(t \bmod 7)$ is used to calculate what day of the week it is at time t. $r^{b,w}_{deliv}$, $r^{p,w}_{deliv}$, and $r^{b,k}_{deliv}$, $r^{p,k}_{delivc}$ denote the power producer's delivery profit on base week, peak week, base weekend, and peak weekend futures, respectively.

The total return received by the power producer from making contractual deliveries is the sum of the profits from the delivery of the six futures:

$$r_{deliv} = r^{b,d}_{deliv} + r^{b,w}_{deliv} + r^{b,k}_{deliv} + r^{p,d}_{deliv} + r^{p,w}_{deliv} + r^{p,k}_{deliv}. \tag{44}$$

3.4.3. Profit from Recovered Margin

The margin previously paid can be recovered once the power producer has completed the delivery of the futures contract. Therefore, this portion of the proceeds can be expressed as follows:

$$r_{recov}^{b,d} = \sum_{i=1}^{l} 24 m_{i,t-i}^{b,d} v_{i,t-i}^{b,d}, \tag{45}$$

$$r_{recov}^{p,d} = \sum_{i=1}^{l} 12 m_{i,t-i}^{p,d} v_{i,t-i}^{p,d}, \tag{46}$$

$$p_{recov}^{b,w} = \sum_{i=1}^{m} \sum_{j=1}^{7m+(t \bmod 7)} 24 m_{i,t-j}^{b,w} v_{i,t-j}^{b,w} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{47}$$

$$p_{recov}^{p,w} = \sum_{i=1}^{m} \sum_{j=1}^{7m+(t \bmod 7)} 12 m_{i,t-j}^{p,w} v_{i,t-j}^{p,w} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{48}$$

$$p_{recov}^{b,k} = \sum_{i=0}^{n} \sum_{j=0}^{7n+(t \bmod 7)} 24 m_{i,t-j}^{b,k} v_{i,t-j}^{b,k} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{49}$$

$$p_{recov}^{p,k} = \sum_{i=0}^{n} \sum_{j=0}^{7n+(t \bmod 7)} 12 m_{i,t-j}^{p,k} v_{i,t-j}^{p,k} \quad \forall \lfloor (t-j+7i)/7 \rfloor = \lfloor t/7 \rfloor, \tag{50}$$

where $r_{recov}^{b,d}$, $r_{recov}^{p,d}$, $r_{recov}^{b,w}$, $r_{recov}^{p,w}$, $r_{recov}^{b,k}$, and $r_{recov}^{p,k}$ denote the power producer's profit from recovered margin on base day, peak day, base week, peak week, base weekend, and peak weekend futures, respectively.

The total return received by the power producer is the sum of the profits from the recovered margin of the six futures:

$$r_{recov} = r_{recov}^{b,d} + r_{recov}^{b,w} + r_{recov}^{b,k} + r_{recov}^{p,d} + r_{recov}^{p,w} + r_{recov}^{p,k}. \tag{51}$$

3.4.4. Profit from Paid Margin

Power producers must pay a margin when entering into a futures contract; therefore, the profit from this portion is negative. It can be expressed as follows:

$$r_{paid}^{d} = \sum_{i=1}^{l} \left(m_{i,t}^{b,d} v_{i,t}^{b,d} + m_{i,t}^{p,d} v_{i,t}^{p,d} \right), \tag{52}$$

$$r_{paid}^{w} = \sum_{i=1}^{m} \left(m_{i,t}^{b,w} v_{i,t}^{b,w} + m_{i,t}^{p,w} v_{i,t}^{p,w} \right), \tag{53}$$

$$r_{paid}^{k} = \sum_{i=0}^{n} \left(m_{i,t}^{b,k} v_{i,t}^{b,k} + m_{i,t}^{p,k} v_{i,t}^{p,k} \right), \tag{54}$$

where r_{paid}^{d}, r_{paid}^{w}, and r_{paid}^{k} denote the margin paid on day, week, and weekend futures, respectively. Add the three to obtain the margin paid by the power producer for entering into a futures contract:

$$r_{paid} = r_{paid}^{d} + r_{paid}^{w} + r_{paid}^{k}. \tag{55}$$

3.4.5. Profit from the Last Time Step

Since the time step of the MADRL process is finite, gains beyond the maximum episode cannot be recovered. Specifically, to recover the margin on futures that have been purchased but for which delivery has not yet been completed, we add that portion of the returned margin to the reward at time T. It is important to clarify that this part of the reward is

virtual; the early return of the margin will not affect the final total reward. Therefore, this portion of the profit can be expressed as follows:

$$r^{d}_{paid} = \sum_{t=T-l}^{T} \sum_{i=1}^{l} \left(m^{b,d}_{i,t} v^{b,d}_{i,t} + m^{p,d}_{i,t} v^{p,d}_{i,t} \right) \qquad \forall i + t > T, \tag{56}$$

$$r^{w}_{paid} = \sum_{t=T-m}^{T} \sum_{i=1}^{m} \left(m^{b,w}_{i,t} v^{b,w}_{i,t} + m^{p,d}_{i,t} v^{p,w}_{i,t} \right) \qquad \forall \lfloor t/7 \rfloor + i > \lfloor T/7 \rfloor, \tag{57}$$

$$r^{k}_{paid} = \sum_{t=T-n}^{T} \sum_{i=0}^{n} \left(m^{b,k}_{i,t} v^{b,k}_{i,t} + m^{p,k}_{i,t} v^{p,k}_{i,t} \right) \qquad \forall \lfloor t/7 \rfloor + i > \lfloor T/7 \rfloor, \tag{58}$$

where r^{d}_{paid},r^{w}_{paid}, and r^{k}_{paid} denote the returns of day, week, and weekend futures in the last time step, respectively. Adding all three components along with the judgment condition yields the total return at the last time step as follows:

$$r_{last} = \begin{cases} r^{d}_{last} + r^{w}_{last} + r^{k}_{last} & t = T \\ 0 & t \neq T \end{cases}. \tag{59}$$

3.4.6. Constraints Penalty

To ensure that the agents' actions adhere to the constraints, a penalty term is incorporated into the reward function. Specifically, $\Gamma_{1,t}$ addresses lower bound constraints, while $\Gamma_{2,t}$ addresses upper bound constraints. These penalty terms can be defined as follows:

$$\Gamma_{1,t} = w_1 \max\{C_t \text{ - } M_{t,max}, 0\} \tag{60}$$

$$\begin{aligned} \Gamma_{2,t} = \omega_2 \Big[& \max\{v^{b,d}_{i,t,min} - v^{b,d}_{i,t}, 0\} + \max\{v^{p,d}_{i,t,min} - v^{p,d}_{i,t}, 0\} \\ & + \max\{v^{b,w}_{i,t,min} - v^{b,w}_{i,t}, 0\} + \max\{v^{p,w}_{i,t,min} - v^{p,w}_{i,t}, 0\} \\ & + \max\{v^{b,k}_{i,t,min} - v^{b,k}_{i,t}, 0\} + \max\{v^{p,k}_{i,t,min} - v^{p,k}_{i,t}, 0\} \Big] \end{aligned} \tag{61}$$

where ω_1 and ω_2 are the penalty term coefficients, respectively.

Therefore, the rewards for each agent at time t can be expressed as follows:

$$r_t = r_{spec} + r_{deliv} + r_{recov} + r_{paid} + r_{last} + \Gamma_{1,t} + \Gamma_{2,t}. \tag{62}$$

By accumulating the rewards received by the agent throughout the cycle T, one obtains its return R as follows:

$$R = \sum_{t=1}^{T} r_t. \tag{63}$$

4. Numerical Experiments

4.1. Simulation of Market Environments

In this study, electricity prices from the Nord Pool market in Germany were used to simulate a realistic market environment [23]. The dataset covers the period from 1 January to 31 December 2022. In addition, German futures trading data were obtained from the Data Sales and Customer Care department of the EEX, which also span the same date range [24]. The base week futures data utilized are shown in Table 2.

Table 2. Sample of base week futures data.

Trading Day	Delivery Start	Delivery End	Lot Size	Settlement Price
2022/1/3	2022/1/10	2022/1/16	168 MWh	216.53 EUR/MWh
2022/1/3	2022/1/17	2022/1/23	168 MWh	215.00 EUR/MWh
2022/1/3	2022/1/24	2022/1/30	168 MWh	229.67 EUR/MWh
2022/1/3	2022/1/31	2022/2/6	168 MWh	240.75 EUR/MWh
2022/1/4	2022/1/10	2022/1/16	168 MWh	215.95 EUR/MWh

Agents generate profit through the purchase, sale, or execution of futures contracts. A comparison of the data in Table 2 with spot electricity prices shows that buying and selling futures contracts serves as a means of profiting from futures price fluctuations. The decision to execute futures contracts is motivated by the potential profit between the futures contract price and the spot electricity price. Therefore, the price relationship between the spot and futures markets significantly influences agent behavior.

Futures contract prices are frequently highly correlated with spot market electricity prices. For instance, Figure 4 illustrates the relationship among base day futures, base week futures, and base weekend futures prices, in conjunction with the average electricity price in the spot market.

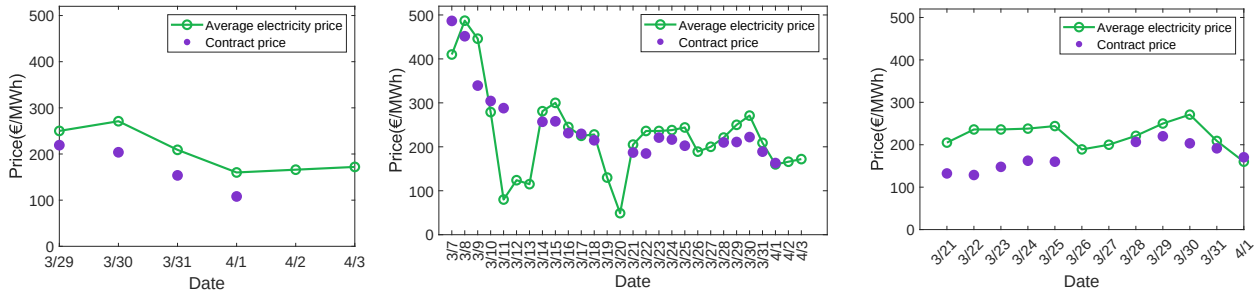


Figure 4. Correlation of day, week, weekend futures prices and average electricity price trends.

Figure 4a illustrates the relationship between the base day futures contract price scheduled for delivery on 4 April and the average electricity price. The base day futures contract can be purchased six days in advance; however, 2 and 3 April are non-business days, resulting in only four days of prices being depicted in the figure. Notably, the futures contract price for April 4 is lower, reflecting the downward trend in the average electricity price in the spot market.

Figure 4b illustrates the relationship between the base week futures contract price scheduled for delivery from 4 April to 10 April and the average electricity price. The base week futures contract can be purchased four weeks in advance. Although the price of the futures contract closely aligns with the spot market price, it exhibits smoother variations.

Figure 4c illustrates the relationship between the base weekend futures contract price, scheduled for delivery on 2 April and 3 April, and the average electricity price. Base weekend futures can be purchased one week prior to the delivery period, as well as on Monday through Friday of that week. Because spot market electricity prices are lower on weekends compared to weekdays, the futures contracts are underpriced relative to spot market prices on the trading day.

Based on the analysis presented, we expanded the dataset by incorporating normally distributed random perturbations into the training data. This approach simulates the complexities of the evolving market environment while preserving the plausibility of futures prices.

We have established the experiment to span 70 days, equivalent to 10 weeks, with $t = 1$ representing Monday, the initial trading day. Day futures contracts have a duration of 6 days, while each week futures contract lasts for 4 weeks and each weekend futures contract lasts for 1 week. The market environment parameters are detailed in Table 3.

Table 3. Market environment parameters.

Parameter	Description	Value
l	Length of day futures contract period	6 days
m	Length of week futures contract period	4 weeks
n	Length of weekend futures contract period	1 week
T	Experiment length	70 days
$M_{t,max}$	Maximum investable amount at time t	10,000 EUR

4.2. Algorithm Design

In the MADRL framework, the traditional MADDPG algorithm is applicable to not only cooperative interactions but also competitive or mixed interactions involving both physical and communicative behaviors [25]. However, it is easy to overestimate the Q values when evaluating an agent's actions. Consequently, improvements have been implemented to address this issue.

The traditional MADDPG algorithm consists of two phases: centralized training and decentralized decision-making.

The centralized training phase is illustrated in Figure 5. MADDPG builds on the Actor-Critic framework, similar to DDPG. This framework includes an Actor neural network that selects actions based on current observations and a Critic neural network that evaluates the value of those actions. The Critic provides feedback in the form of the actual Q-value associated with the current action-observation pair. Consequently, the Actor is trained using this feedback, while the Critic refines its estimates by comparing its predicted Q-value to the actual Q-value. MADDPG employs a centralized training approach, in contrast to DDPG's single-agent model. During training, each agent's Critic network learns not only from its own Actor's actions but also from the actions of other agents' Actors. This is accomplished through a replay buffer that stores actions and feedback—such as observations and rewards—from all agents. These shared experiences can be utilized to enhance the training of each agent's Critic network.

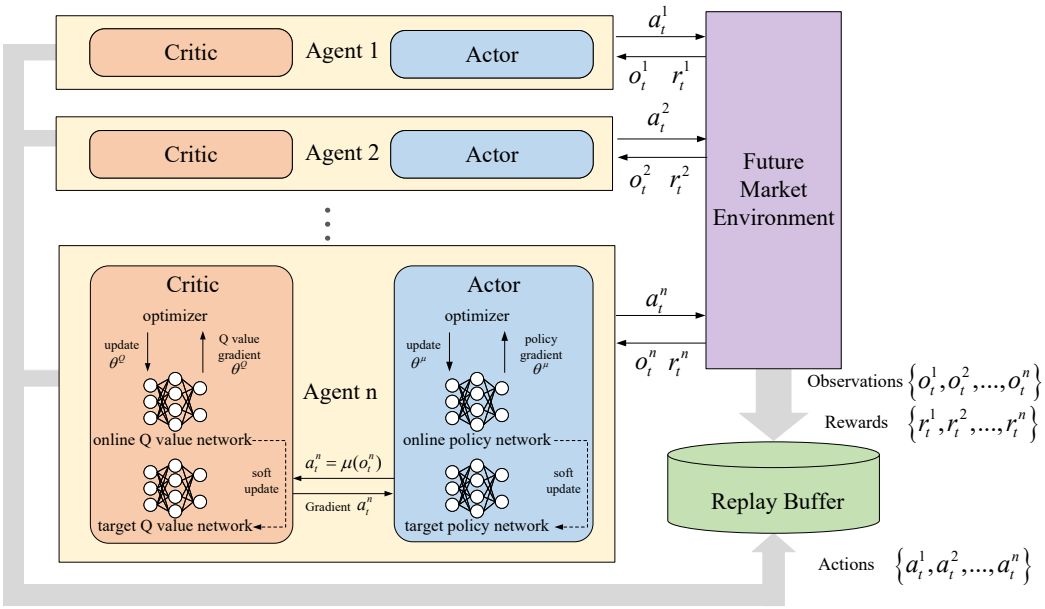


Figure 5. The centralized training phase.

The decentralized decision-making phase is illustrated in Figure 6. During this phase, each agent relies solely on its Actor network for decision-making, without feedback from the Critic. Each agent's Actor network selects actions based on its own state, operating independently of the states and actions of other agents.

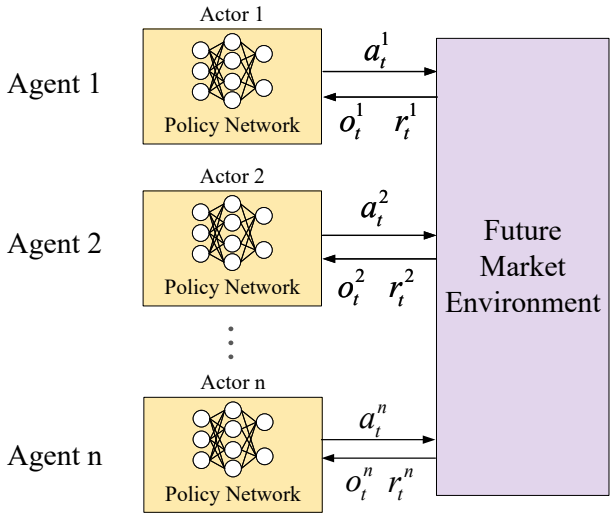


Figure 6. The decentralized decision-making phase.

Given that electricity is an energy source that necessitates the prompt balancing of supply and demand, its price fluctuations tend to be more pronounced in comparison to other energy products, such as oil and natural gas. The volatility of electricity prices gives rise to a dual fluctuation in both electricity spot prices and futures prices. In this paper, we utilize the profit generated by power producers in the electricity financial market as a reward function for deep reinforcement learning. The elevated volatility of the market, in turn, tends to result in disproportionate returns within a relatively short period of time.

In the centralized training phase, this can readily result in the overestimation of the Q-value associated with the action taken by the intelligence during the training process of the deep reinforcement learning algorithm.

Consequently, we devised a Dual-Q MADDPG algorithm comprising two critic objective networks, with the minimum of the outputs of the two networks being taken as the Q-value. This effectively addresses the issue of overestimation in deep reinforcement learning. From another perspective, this approach reduces the sensitivity of the critic target network to price fluctuations, thereby facilitating the attainment of higher overall returns over a long time horizon. The specific improvements are illustrated in Figure 7.

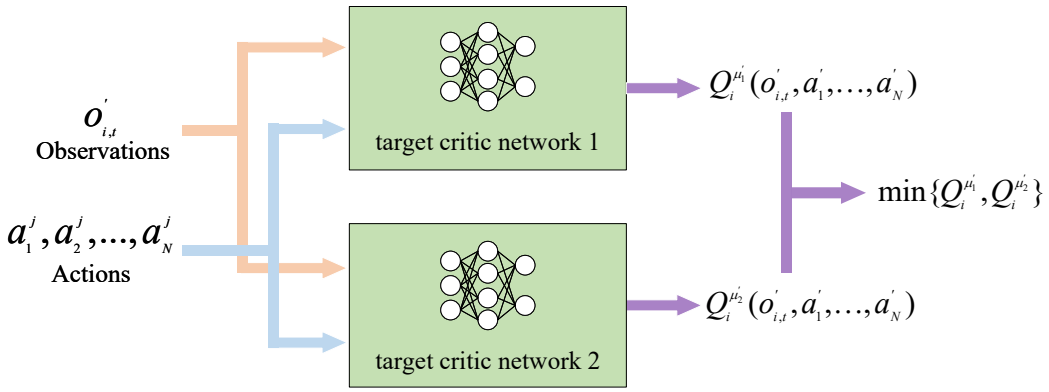


Figure 7. The dual-Q critic network.

The improved MADDPG algorithm is referred to as Dual-Q MADDPG. The flow of the Dual-Q MADDPG algorithm is presented in the pseudocode Algorithm 1.

Algorithm 1 Algorithm of dual-Q MADDPG

1: **for** episode = 1 to M **do**
2: Set the initial observation for each agent i, including the initial price and margin for each type of futures contract.
3: **for** time $t = 1$ to T **do**
4: For each agent i, determine the contracts to buy by calculating the optimal number of futures based on market conditions and risk preferences: $a_i = \boldsymbol{\mu}_{\theta_i}(o_i) + \mathcal{N}_t$.
5: Purchase the appropriate volumes of futures, receive rewards r_t, and transition to the next observation $o'_{i,t}$.
6: Store $(o_{i,t}, a_{i,t}, r_t, o'_{i,t})$ in the replay buffer $\mathcal{D}$.
7: Update observation: $o_{i,t} \leftarrow o'_{i,t}$.
8: **for** agent $i = 1$ to $\mathcal{N}$ **do**
9: Sample a random minibatch of $\mathcal{S}$ samples $(o^j_{i,t}, a^j_{i,t}, r^j_t, o'^j_{i,t})$ from $\mathcal{D}$.
10: Set $y^j = r^j_i + \gamma \min(Q_i^{\mu'_1}(o'_{i,t}, a'_1, \ldots, a'_N), Q_i^{\mu'_2}(o'_{i,t}, a'_1, \ldots, a'_N))$.
11: Update the online Critic network by minimizing loss: $\mathcal{L}(\theta_i) = \frac{1}{S}\sum_j \left(y^j - Q_i^{\boldsymbol{\mu}}(o'_{i,t}, a^j_1, \ldots, a^j_N)\right)^2$.
12: Update the Actor network using the policy gradient:
13: $\nabla_{\theta_i} J \approx \frac{1}{S}\sum_j \nabla_{\theta_i}\boldsymbol{\mu}_i(o^j_i)\nabla_{a_i} Q_i^{\boldsymbol{\mu}}(o'_{i,t}, a^j_1, \ldots, a_i, \ldots, a^j_N)\big|_{a_i=\boldsymbol{\mu}_i(o^j_i)}$.
14: **end for**
15: Soft update target network parameters for each agent i:
16: $\theta'_i \leftarrow \tau\theta_i + (1-\tau)\theta'_i$.
17: **end for**
18: **end for**

In line 10 of the pseudocode above, the traditional MADDPG algorithm updates the critic's online network using a single target network Q value. In the proposed Dual-Q MADDPG, the minimum of the two target network Q values is utilized for the update. This approach mitigates the issue of potential overestimation of the Q values.

During the training of the Daul-Q MADDPG algorithm, the primary parameters are set as indicated in Table 4.

Table 4. Algorithm parameters.

Parameter	Value	Description
Discount rate	0.97	Factor for discounting future rewards.
Max episode length	70	Maximum duration (in time steps) of a single episode.
Total episodes	2000	Total number of training episodes.
Learning rate (actor)	0.0001	Learning rate for the actor's neural network.
Learning rate (critic)	0.001	Learning rate for the critic's neural network.
Noise rate	0.1	Rate of noise added during sampling for exploration.
Soft update weighting	0.01	Weighting factor between target and online networks.
Buffer size	500,000	Size of the replay buffer for storing experiences.
Batch size	256	Number of samples used in each training batch.
Evaluation frequency	100	Model evaluation is performed every 100 episodes.

4.3. Experiment Results

4.3.1. Return Evolution During Trainin

The MADDPG algorithm is implemented in Python using the PyTorch library. The experiment was conducted on a computer equipped with an Intel Core i7-11800H processor running at 2.30 GHz. The training process involved 700,000 steps and lasted 5.5 h.

During the training of MADDPG, we tested the model every 100 episodes, averaging the returns of five episodes for each test. In total, 200 model tests were conducted over 2000 episodes. After applying a sliding average, the comparison of returns obtained using the traditional MADDPG algorithm and the Dual-Q MADDPG algorithm is shown in Figure 8.

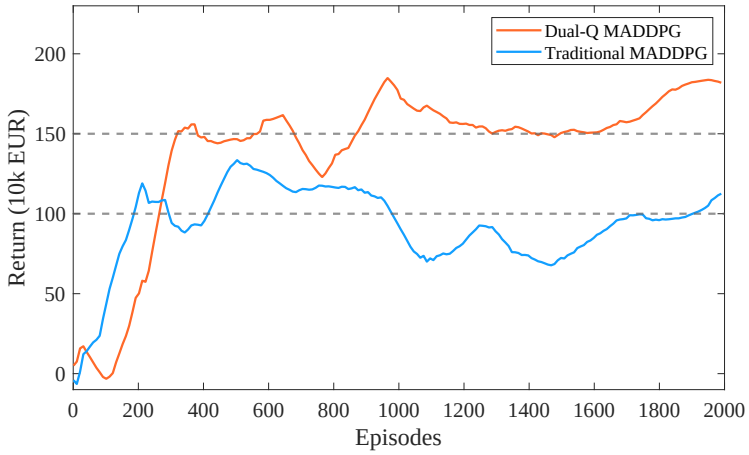


Figure 8. Comparison of the returns of the two algorithms.

From Figure 8, it is evident that compared to the traditional MADDPG algorithm, the improved Dual-Q MADDPG algorithm achieves higher returns after convergence, although its convergence speed is slower in the initial stages. This algorithm enables the power producer to realize a profit of approximately 1500 k EUR, while the conventional MADDPG algorithm yields a profit of around 1000 k EUR.

4.3.2. Behavioral Patterns Through Actions

To apply the Dual-Q MADDPG algorithm to a specific power producer, the model must first be trained using historical data. Once the model reaches maturity, indicated by stable agent returns, it can be implemented. The trained agents will then make decisions based on the current state of the power producer. Finally, to develop a complete portfolio strategy, the actions of multiple agents must be combined.

The behavioral strategies employed by the agents are analyzed using the base week futures contract, which began delivery on 4 April 2022, as an illustrative example. This contract was listed on 7 March 2022, and the strategies used by Agent 2 over the 20 trading days are shown in Figure 9.

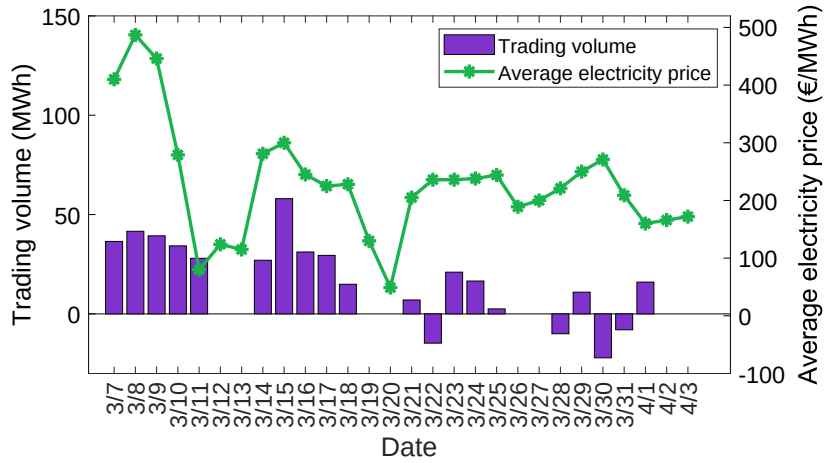


Figure 9. Agent 2's actions on the base week futures contract.

An analysis of Agent 2's behavior, as shown in Figure 9, indicates that minimal futures are sold when the spot price volatility is high compared to the futures price. This is because the market environment at that time is characterized by high volatility and uncertainty. Selling too many futures could lead to an insufficient number of contracts available for hedging price risk during the delivery period. Conversely, when prices are relatively stable, Agent 2 engages in more speculative trading, buying at low prices and selling at high prices based on market forecasts. The other agents follow a similar strategy.

4.3.3. Status of Contract Implementation

We further analyze the performance of the power producer's investments in various futures contract and calculate the actual amount of power delivered by the power producer on each day. The delivery status of day futures, week futures, and weekend futures is shown in Figure 10:

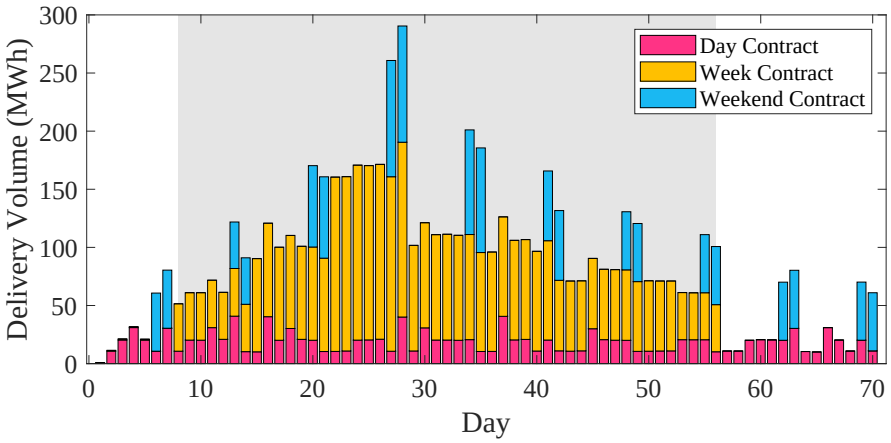


Figure 10. Daily delivery volumes for day, week, and weekend contracts.

As shown in Figure 10, the daily delivery volume of the futures contract fluctuates around 10 MWh. The delivery volume of the weekend futures remains relatively stable at the end of each week, with approximately 10 MWh of electricity delivered per hour on Saturdays and Sundays. Cycle futures serve as the primary tool for balancing significant price fluctuations. The week contract delivery volume varies over time. From the 8th to the 57th day, spanning the second to the eighth week, large quantities of cyclical goods are delivered daily.

To better understand the trading strategy of the agents, we analyze it along with fluctuations in electricity prices in the spot market. The average price of electricity during the base period and the peak hours are shown in Figure 11.

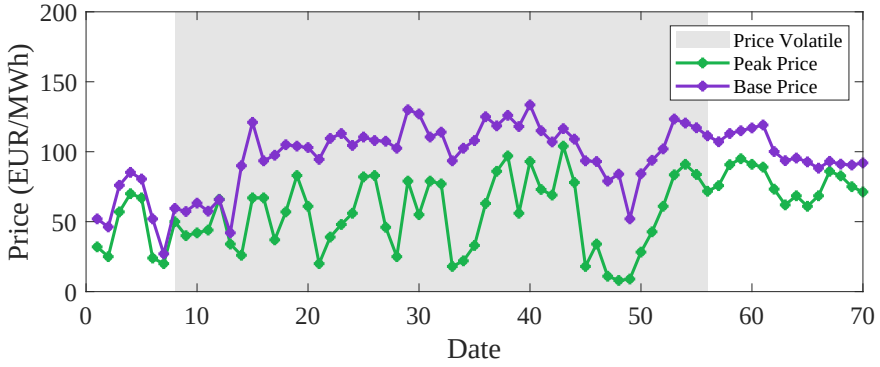


Figure 11. Base day and peak spot market electricity prices.

As illustrated in Figure 11, the average intraday electricity price in the spot market experiences significant fluctuations between days 7 and 58. During this period, agents anticipate future conditions and choose to hedge against price risk by purchasing a substantial number of futures contracts in advance. Following day 58, the volatility of the spot market price diminishes, and there are no futures deliveries.

Further analysis of the delivery of base and peak futures reveals the daily contract deliveries, as depicted in Figure 12.

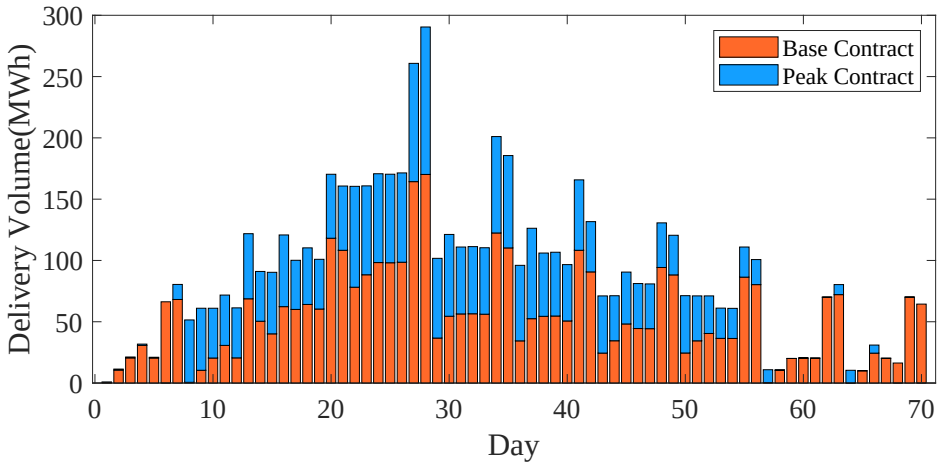


Figure 12. Daily delivery volumes of the base and peak contracts.

From Figure 12, it is evident that the delivery volume of the base futures contract consistently exceeds that of the peak futures contract. Additionally, when spot market prices are more volatile, the delivery volume of the peak futures contract increases. Notably, the delivery volume of the peak futures remains relatively low from day 57 to day 70, corresponding to the 9th and 10th weeks.

4.3.4. Analysis of Episode Length

In DRL, each Markov decision process is modeled with a different maximum round length and trained to yield disparate results. Consequently, we have augmented the original simulation of 70-day trading cycles with 105-day versus 140-day returns. The original Dual-Q MADDPG algorithm is trained to generate the return curve depicted in Figure 13.

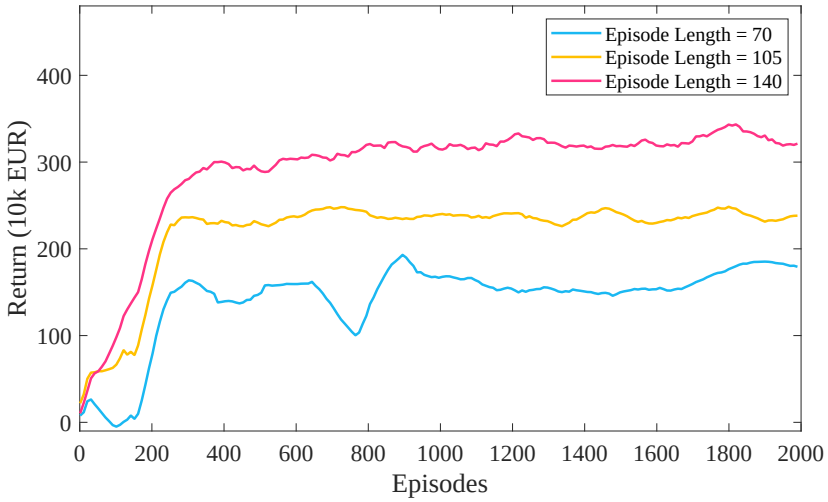


Figure 13. Comparison of the returns of the three episode length.

As shown in Figure 13, increasing the episode length leads to a higher total return and improved training stability. Specifically, with an episode length of 70, the average return after stabilizing is approximately 1.5 million EUR, whereas with an episode length of 140, the average return increases to about 3 million EUR.

To further analyze the average daily gain, the trained and matured model was tested under various episode lengths. The results, which include both model training time and a comparison with the traditional MADDPG algorithm, are presented in Table 5.

Table 5. Comparison of different algorithms.

Algorithm	Episode Length	Average Daily Profit	Training Time
MADDPG	70 days	15.67 k EUR	289 min
	105 days	16.83 k EUR	595 min
	140 days	16.39 k EUR	976 min
Daul-Q MADDPG	70 days	21.44 k EUR	327 min
	105 days	22.36 k EUR	897 min
	140 days	22.61 k EUR	1134 min

Table 5 leads to two key conclusions. First, the average daily profit obtained using different algorithms varies significantly. The Dual-Q MADDPG algorithm proposed in this paper shows better adaptability to the highly volatile electricity futures market, resulting in higher returns. Second, while the average daily profit across different episode lengths—i.e., simulation time durations—is similar, the variation in model training time is more significant. This may be limited by the computational resources available during the experiments. Therefore, when applying the algorithm in practice, adjustments should be made based on the actual computing resources to accommodate different environments.

5. Conclusions

This study addresses the previously unexplored area of providing guidance to power generation companies in navigating the complexities of electricity futures markets. The conventional MADDPG algorithm has been enhanced to account for the high level of unpredictability inherent in these markets. Numerical experiments demonstrate that the model performs effectively in the Nordic power futures market, enabling power producers to achieve better returns from their futures trading activities. This research contributes to lowering barriers to participation in the electricity financial market, thereby promoting the broader use of electricity financial instruments and enhancing market liquidity and stability.

However, there are opportunities for further development. First, in considering futures trading strategies, it is crucial to prioritize risk management alongside the pursuit of optimal returns, as market participants typically aim for higher returns within a specified risk tolerance. Second, modern portfolio theory suggests that systematic market risk can be reduced by diversifying asset allocations. Therefore, when constructing a multi-asset portfolio, careful attention must be paid to the proportionate allocation of assets.

Additionally, this paper focuses on power producers. Future studies could also explore the perspectives of other market participants, such as financial institutions and large-scale electricity consumers in the electricity financial market.

Author Contributions: Conceptualization, Y.W. and E.S.; methodology, E.S. and Y.W.; software, E.S. and Y.W.; validation, Y.W. and E.S.; formal analysis, E.S. and Y.W.; investigation, E.S. and Y.W.; resources, Y.W.; data curation, E.S. and Y.W.; writing—original draft preparation, E.S.; writing—review and editing, E.S.; visualization, Y.X., J.H. and C.F.; supervision, Y.X., J.H. and C.F.; project administration, Y.X., J.H. and C.F.; funding acquisition, Y.W., Y.X., J.H. and C.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Natural Science Foundation of China under Grant 52107129.

Data Availability Statement: The datasets presented in this article are not readily available due to time limitations. Requests to access the datasets should be directed to 202105031119@zjut.edu.cn

Conflicts of Interest: Author Yizheng Wang was employed by the company Zhejiang Electric Power Company. Authors Yang Xu and Jiahua Hu were employed by the company State Grid Zhejiang Electric Power Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EEX	European Energy Exchange
DRL	Deep reinforcement learning
MADPL	Multi-agent deep reinforcement learning
MADDPG	Multi-agent deep deterministic policy gradient
ICE	Intercontinental Exchange
RL	Reinforcement learning
DL	Deep learning
Dec-POMDP	Decentralized partially observable Markov decision process
CTDE	Centralized training with distributed execution

References

1. Deng, S.-J.; Oren, S.S. Electricity derivatives and risk management. *Energy* **2006**, *31*, 940–953. [CrossRef]
2. Mork, E. Emergence of financial markets for electricity: A European perspective. *Energy Policy* **2001**, *29*, 7–15. [CrossRef]
3. Tsaousoglou, G.; Giraldo, J.S.; Paterakis, N.G. Market mechanisms for local electricity markets: A review of models, solution concepts and algorithmic techniques. *Renew. Sustain. Energy Rev.* **2022**, *156*, 111890. [CrossRef]
4. Liu, C.; Chen, X.; Hua, H.; Yu, K.; Jiang, Y.; Yin, M. Portfolio Strategy of Power Producer Considering Energy Storage in Spot Market, Ancillary Market and Option Market. In Proceedings of the 2022 IEEE International Conference on Energy Internet (ICEI), Stavanger, Norway, 28–29 December 2022; pp. 13–18.
5. Yucekaya, A. Electricity trading for coal-fired power plants in Turkish power market considering uncertainty in spot, derivatives and bilateral contract market. *Renew. Sustain. Energy Rev.* **2022**, *159*, 112189. [CrossRef]
6. Sheybani, H.R.; Buygi, M.O. Put option pricing and its effects on day-ahead electricity markets. *IEEE Syst. J.* **2017**, *12*, 2821–2831. [CrossRef]
7. Jaeck, E.; Lautier, D. Volatility in electricity derivative markets: The Samuelson effect revisited. *Energy Econ.* **2016**, *59*, 300–313. [CrossRef]
8. Liedes, T. Power Derivatives Market Trend Prediction with Machine Learning and Technical Analysis. Bachelor's Thesis, Tampere University, Tampere, Finland, 2023.
9. Deng, S.; Zhu, Y.; Huang, X.; Duan, S.; Fu, Z. High-frequency direction forecasting of the futures market using a machine-learning-based method. *Future Internet* **2022**, *14*, 180. [CrossRef]
10. Creamer, G.G.; Lee, C. A multivariate distance nonlinear causality test based on partial distance correlation: A machine learning application to energy futures. In *Machine Learning and AI in Finance*; Routledge: Oxfordshire, UK, 2021; pp. 82–93.
11. Saâdaoui, F.; Mefteh-Wali, S.; Jabeur, S.B. Multiresolutional statistical machine learning for testing interdependence of power markets: A Variational Mode Decomposition-based approach. *Expert Syst. Appl.* **2022**, *208*, 118161. [CrossRef]
12. Kawy, R.A.; Abdelmoez, W.M.; Shoukry, A. Financial portfolio construction for Quantitative Trading using Deep learning technique. In Proceedings of the International Conference on Artificial Intelligence and Soft Computing, Online, 21–23 June 2021; pp. 3–14.
13. Zhang, H.; Jiang, Z.; Su, J. A deep deterministic policy gradient-based strategy for stocks portfolio management. In Proceedings of the 2021 IEEE 6th International Conference on Big Data Analytics (ICBDA), Virtual, 5–8 March 2021; pp. 230–238.
14. Pham, U.; Luu, Q.; Tran, H. Multi-agent reinforcement learning approach for hedging portfolio problem. *Soft Comput.* **2021**, *25*, 7877–7885. [CrossRef] [PubMed]
15. Liang, Z.; Chen, H.; Zhu, J.; Jiang, K.; Li, Y. Adversarial deep reinforcement learning in portfolio management. *arXiv* **2018**, arXiv:1808.09940.
16. Liu, X.-Y.; Yang, H.; Chen, Q.; Zhang, R.; Yang, L.; Xiao, B.; Wang, C.D. FinRL: A deep reinforcement learning library for automated stock trading in quantitative finance. *arXiv* **2020**, arXiv:2011.09607. [CrossRef]
17. European Energy Exchange (EEX). Available online: https://www.eex.com/en/markets/power (accessed on 25 September 2024).
18. Intercontinental Exchange (ICE). Available online: https://www.ice.com/products/57609943/German-Power-Financial-Peak-Daily-Futures (accessed on 25 September 2024).
19. Yu, X.; Li, Y.; Lu, J.; Shen, X. Futures hedging in crude oil markets: A trade-off between risk and return. *Resour. Policy* **2023**, *80*, 103147. [CrossRef]

20. Xiao, J.; Wang, Y. Macroeconomic uncertainty, speculation, and energy futures returns: Evidence from a quantile regression. *Energy* **2022**, *241*, 122517. [CrossRef]
21. Arulkumaran, K.; Deisenroth, M.P.; Brundage, M.; Bharath, A.A. Deep reinforcement learning: A brief survey. *IEEE Signal Process. Mag.* **2017**, *34*, 26–38. [CrossRef]
22. Li, S.E. Deep reinforcement learning. In *Reinforcement Learning for Sequential Decision and Optimal Control*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 365–402.
23. Nord Pool. Available online: https://data.nordpoolgroup.com/auction/day-ahead/prices (accessed on 25 September 2024).
24. European Energy Exchange (EEX) Futures Market Data. Available online: https://www.eex.com/en/market-data/power/futures (accessed on 25 September 2024).
25. Lowe, R.; Wu, Y.I.; Tamar, A.; Harb, J.; Pieter Abbeel, O.; Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. *arXiv* **2017**, arXiv:1706.02275.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.