sustainability

Article

Deep-Reinforcement-Learning-Based Vehicle-to-Grid Operation Strategies for Managing Solar Power Generation Forecast Errors

Moon-Jong Jang [1] and Eunsung Oh [2,*]

[1] Smart Power Distribution Laboratory of Korea Electric Power Research Institute, Korea Electric Power Corporation, Daejeon 34056, South Chungcheong, Republic of Korea; m.jang@kepco.co.kr
[2] Department of Electrical Engineering, College of IT Convergence, Global Campus, Gachon University, Seongnam-si 13120, Gyeonggi-do, Republic of Korea
[*] Correspondence: esoh@gachon.ac.kr; Tel.: +82-31-750-5346

Abstract: This study proposes a deep-reinforcement-learning (DRL)-based vehicle-to-grid (V2G) operation strategy that focuses on the dynamic integration of charging station (CS) status to refine solar power generation (SPG) forecasts. To address the variability in solar energy and CS status, this study proposes a novel approach by formulating the V2G operation as a Markov decision process and leveraging DRL to adaptively manage SPG forecast errors. Utilizing real-world data from the Korea Southern Power Corporation, the effectiveness of this strategy in enhancing SPG forecasts is proven using the PyTorch framework. The results demonstrate a significant reduction in the mean squared error by 40% to 56% compared to scenarios without V2G. Our investigation into the effects of blocking probability thresholds and discount factors revealed insights into the optimal V2G system performance, suggesting a balance between immediate operational needs and long-term strategic objectives. The findings highlight the possibility of using DRL-based strategies to achieve more reliable and efficient renewable energy integration in power grids, marking a significant step forward in smart grid optimization.

Keywords: blocking probability; charging station; deep reinforcement learning; electric vehicle; forecast error; power generation forecasting; reinforcement learning; solar; vehicle-to-grid operation

Citation: Jang, M.-J.; Oh, E. Deep-Reinforcement-Learning-Based Vehicle-to-Grid Operation Strategies for Managing Solar Power Generation Forecast Errors. *Sustainability* **2024**, *16*, 3851. https://doi.org/10.3390/su16093851

Academic Editors: Ghulam Abbas, Irfan Ahmad Khan, Mousa Marzband and Muhammad Faisal Nadeem Khan

Received: 12 April 2024
Revised: 30 April 2024
Accepted: 2 May 2024
Published: 3 May 2024

1. Introduction

1.1. Motivation

Power generation from renewable sources has undergone an unprecedented increase. The International Energy Agency (IEA) reports that annual additions to renewable capacity will surge by 50% year on year in 2023, reaching approximately 510 GW, and the capacity is projected to increase to 7300 GW by 2028 [1]. Notably, solar power has become the predominant form of renewable energy, contributing to over 70% of the additions globally. The preference for solar energy stems from its dual benefits: addressing environmental problems and offering economic advantages through a lower levelized cost of energy (LCOE) compared to traditional fossil fuels in developed nations [2].

However, the assimilation of solar power into the grid poses distinct challenges, primarily because of its variable nature. Solar power generation (SPG) is subject to natural phenomena, such as sunlight intensity and duration, causing daily and seasonal fluctuations [3]. Consequently, the precise forecasting of SPG has become vital for maintaining grid stability and optimizing the use of solar energy. Despite technological advancements, forecasting inaccuracies remain a major challenge. These inaccuracies, which are caused by unpredictable weather changes and the inherent limitations of forecasting models, require grid operators to devise robust strategies to use forecasted outputs in actual generation [4].

Alongside the growth in renewable energy, electric vehicles (EVs) are gaining traction owing to environmental sustainability concerns. In 2022, EVs represented 14% of all new car sales, and this figure is expected to increase to approximately 35% by 2030 [5]. The increasing prevalence of EVs has opened up new avenues for their utilization as adaptable energy resources [6]. Owing to their rapid response capabilities, EVs are ideal for integration into vehicle-to-grid (V2G) technology [7]. This innovative application allows for a bi-directional flow of energy, enabling EVs to either draw power from, or supply power back to, the grid, as required, thereby acting as a dynamic energy storage solution. The flexibility offered by EVs can be strategically employed to address the challenges posed by SPG forecast errors, demonstrating the potential of V2G technology to enhance the resilience and efficiency of solar energy integration into the grid.

1.2. Contributions

This study contributes to the development and validation of a DRL-based V2G operation strategy that meticulously considers the status of CS within its framework. This strategy is specifically designed to manage the uncertainties associated with SPG forecast errors, highlighting a critical area in renewable energy integration, where predictive inaccuracies can significantly impact grid operations. The main contribution can be summarized as follows:

- Charging-station-centric V2G operation modeling: Acknowledging the dynamic interplay between EV charging and discharging activities and the operational status of the CS, this study introduces an approach that seeks to optimize V2G operations. By considering the availability and capacity of the CS and the potential increase in service time resulting from V2G actions, we addressed the operational challenge of maintaining efficient CS utilization while managing SPG forecast error uncertainties. This was achieved by formulating the V2G operation problem as a Markov decision process (MDP), where sequential decision making was employed to identify the optimal charging or discharging actions of EVs in the context of CS utilization.
- Enhanced DRL model with advanced learning techniques: To ensure the practical applicability and effectiveness of the proposed strategy, we incorporated a penalty-based reward system into the DRL model. This was complemented by advanced techniques, such as the use of a replay buffer and a target network with delayed updates, to enhance the learning efficiency and performance of the model. The feasibility and performance of the proposed DRL-based V2G operation strategy were rigorously validated through experimental studies using a real CS status measurement dataset from Korea. The present analysis provides a comprehensive evaluation of the impact of this strategy on grid management and demonstrates its potential to mitigate the adverse effects of SPG forecast errors.
- Insights and implications of V2G operation: By integrating V2G operations with intelligent grid management practices, this study highlights the significance of leveraging advanced DRL methodologies to address the complexities and uncertainties inherent in renewable energy systems. This contribution advances state-of-the-art V2G operation strategies and offers valuable insights for future developments in renewable energy integration and grid optimization.

The remainder of this paper is structured as follows. Section 2 reviews prior works and discusses the research gap. Section 3 details the SPG model within a V2G system context and outlines the specific V2G operation problem addressed in this study. Section 4 describes the methodology behind the proposed V2G operation strategy and provides insights into the design process and the rationale behind the chosen approach. In Section 5, we present measurement studies and discuss the application and effectiveness of the proposed strategy, highlighting the key findings and implications. Finally, Section 6 concludes the paper and summarizes the contributions of the study, implications for future V2G operations, and potential directions for further research.

2. Literature Review

2.1. Prior Works

To address the uncertainty in SPG forecasting, scholars have approached the problem from several perspectives, including resource dispatch from a utility perspective and demand response from the demand side. Ferris and Philpott explored the optimization of investments in renewable energy generation within an electrical system dominated by hydro power, particularly under the challenge of inflow uncertainty, which threatens energy shortages [8]. Their work integrated uncertain seasonal hydroelectric supply and short-term renewable supply variability into a two-stage stochastic programming framework, considering hydroelectric and battery storage solutions. Gheouany et al. proposed a multi-stage energy management system for microgrids, incorporating a multi-objective particle swarm optimization algorithm for model predictive control and reactive layers using extremum seeking for real-time optimization [9]. The system effectively addressed forecast uncertainty and ensures significant reductions in daily energy costs and battery storage system degradation. Srinivasan et al. developed a cost-optimization model that accommodated load and generation uncertainty by engaging in electrical ancillary service markets [10]. By aiming for decarbonization, their model attempted to design a multi-energy system in a cost-optimized manner while providing flexibility through battery storage after addressing weather-related forecasting errors. Bodong et al. introduced an autoregressive moving-average probabilistic model for economic management and planning that integrated renewable energy sources and a battery-based demand-side management program within a multi-energy market context [11]. A hybrid algorithm combining the seagull optimization algorithm and a genetic algorithm was developed to effectively manage renewable energy production uncertainties. Zheng et al. proposed a strategy to enhance the management of demand response in renewable energy systems by combining deep-learning-based short-term renewable power generation prediction and fuzzy-logic-based energy storage system operation [12]. Battery and supercapacitor system operation has been proposed to improve the stability and reliability of renewable energy systems.

In existing studies on managing SPG forecast uncertainty, battery energy storage systems (BESSs) serve as flexibility resources because of their controllability and designation as fixed energy resources, thereby eliminating uncertainty in their flexibility role. However, BESSs still present an economic barrier owing to their high cost and limitations in terms of installation and mobility [13]. This cost barrier encompasses the initial capital expenditure and maintenance and potential replacement expenses over the system's lifetime [14]. Additionally, the substantial size and weight of these systems can complicate their deployment in areas with limited space or remote locations, thereby affecting the scalability and flexibility of energy storage solutions.

The research on V2G is advancing, utilizing EVs as flexible resources without requiring additional investment costs. Unlike BESSs, EVs function as dynamic resources because their availability and capacity for grid services vary based on the EV owners' preferences. Consequently, much of the research involving EVs as flexible resources has been dedicated to developing methodologies for V2G operations. Alfaverh et al. investigated the integration of EVs into a grid for supplementary frequency regulation (SFR) through V2G technology, presenting an optimal V2G control strategy that employed deep reinforcement learning (DRL), specifically the deep deterministic policy gradient (DDPG) algorithm, for the dynamic adjustment of V2G power scheduling [15]. This study utilized randomly generated models for the EV status. Maeng et al. introduced a DRL approach for managing EV charging and discharging to mitigate peak loads, highlighting the capacity of V2G technology to enhance grid stability by enabling bidirectional energy flow, thereby enabling EVs to supply excess energy back to the grid during peak periods [16]. In addition to bolstering grid reliability, it offered revenue generation opportunities for EV owners, considering a probabilistic-distributed EV status. Dong et al. explored optimizing the use of EVs within intelligent transportation systems (ITSs) to diminish peak loads, underscoring the dual functionality of EVs as both a transportation method and a distributed energy resource (DER) for efficient peak load management [17]. Their approach is notable for permitting EVs, viewed as agents, to autonomously decide on energy discharge times back into the grid with optimization through centralized training and local execution, thus maintaining individual EV owners' autonomy and privacy. However, the study did not model the EV status. Wang et al. optimized EV charging station (CS) scheduling to improve the operational efficiency and economic benefits, acknowledging the dynamic and real-time alignment between EVs' charging demands and CS resources such as DERs [18]. This research framed the CS scheduling problem as a finite MDP and integrated a DRL method using a modified rainbow algorithm to devise a time-scale-based CS scheduling scheme that accommodated mismatches in energy scale resources. The model assumed a uniformly random EV status. Shibl et al. presented a DRL-based power system management solution for EV charging management considering the impact of fast charging, conventional charging, and V2G operations [19]. The study balanced EV users' needs with utility demands, ensuring grid stability and minimizing adverse effects such as voltage fluctuations, power losses, and overloads from EV charging. This research did not presuppose EV status but demonstrated outcomes based on a case sensitivity analysis.

In examining the landscape of existing research, it is clear that a significant proportion of studies on V2G operations rely on RL approaches. The preference for RL methods is primarily owing to their model-free characteristic, which is particularly well suited for navigating and modeling the inherently dynamic nature of V2G environments. This capacity allows for adaptive learning in situations where the environmental variables and system dynamics are not fully known or are too complex to model explicitly.

Despite the adaptability of RL methods, a notable limitation observed across many studies is the simplistic treatment of the EV status. Often, the EV status is considered either as a random variable or assumed to be predefined. It overlooks the variable behavior of EVs within the V2G system, such as their availability for charging or discharging at any given time, which is a critical factor in the application of V2G technologies. In the context of V2G operations, the functionality and efficiency are significantly influenced by the operational strategies for charging the CS. These stations serve as the nexus for the interaction between the grid and EVs, where the scheduling and management of EV charging and discharging are crucial. Although some research [18] delves into CS scheduling problems, it often limits this exploration to viewing EV resources through the lens of an aggregator's role. This perspective does not fully address the dynamic interplay between EV charging and discharging actions and their impact on CS status. Reliable V2G operations require an approach that fully considers the impact of EV charging and discharging activities on CS status.

2.2. Research Gaps

While existing studies have significantly advanced the integration of renewable energy resources into grid systems and explored the utility of V2G technology, several gaps remain, particularly in the context of optimizing these systems:

- Dynamic Integration: Most studies, such as those by Ferris and Philpott, Gheouany et al., and others, focus on optimizing renewable energy systems with a static or quasi-static approach to resource management. There is a lack of emphasis on dynamic integration where real-time data and continuous learning from the environment can profoundly impact operational strategies. Our study addressed this by implementing a DRL-based strategy that continuously adapted to changing conditions, enhancing the responsiveness of V2G systems to SPG forecast errors.

- Comprehensive V2G Operation Modeling: Although research such as that by Alfaverh et al. and Wang et al. has begun to explore the use of DRL in V2G operations, these studies often do not fully incorporate the complex interdependencies between EV charging demands, CS operational status, and grid needs. Our approach extended this by formulating the V2G operation as an MDP that captured a broader range of variables and scenarios, providing a more holistic view of V2G dynamics.

By addressing these gaps, our study aimed to significantly advance the field of smart grid management, offering innovative solutions for the integration of renewable energy sources through sophisticated DRL-based V2G operations.

3. System Model and Problem Formulation

3.1. SPG Model with a V2G System

In this study, a grid-connected SPG system was considered, as shown in Figure 1.

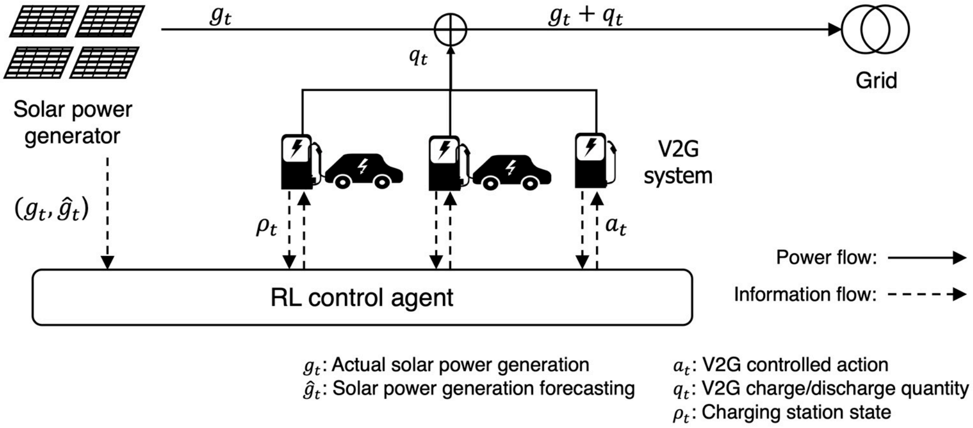

g_t
$g_t + q_t$
q_t
Grid
Solar power generator
V2G system
$(g_t, \hat{g}_t)$
ρ_t
a_t
Power flow:
Information flow:
RL control agent
g_t: Actual solar power generation
$\hat{g}_t$: Solar power generation forecasting
a_t: V2G controlled action
q_t: V2G charge/discharge quantity
ρ_t: Charging station state

Figure 1. SPG system model with a V2G system.

3.1.1. SPG Forecast Error Model

Let g_t and $\hat{g}_t$ denote the actual SPG and its forecast at time t, respectively. The SPG forecast error at time t was defined as follows:

$$e_t = g_t - \hat{g}_t. \tag{1}$$

A V2G system was added to manage the SPG forecast error, as shown in Figure 1. The V2G charging or discharging energy q_t compensated for the SPG forecast error, as follows:

$$\epsilon_t = e_t - q_t. \tag{2}$$

The objective of the V2G charging or discharging operation was to eliminate the SPG forecast error in Equation (2). Therefore, a positive action by q_t expressed the charging operation, and vice versa.

3.1.2. V2G Model as Flexibility Resource

V2G systems are designed to harness parked EVs as flexible resources. This concept revolves around using EV batteries effectively as a buffer to address the discrepancies between the forecasted and actual power demand. Flexibility resources refer to the capability to supply energy that can be readily dispatched to compensate for SPG forecast errors. Therefore, modeling the V2G system as a flexible resource primarily focuses on the characteristics and capabilities of EV CSs.

At decision time t, if no EVs are present at the CS, it becomes impossible to provide a flexible resource. However, if EVs are present at the CS EV, then it is assumed that all

EVs can be utilized as a flexibility resource. Although the capacity available for flexibility can be estimated based on preferences and associated probabilities, the capacity available at decision time t is determined as a deterministic value. This means that while the potential to contribute to the grid's flexibility can vary based on the likelihood of EV presence and the owners' willingness to participate in V2G services, the actual capacity that can be deployed at any given moment is fixed and can be precisely calculated. This deterministic approach promotes more accurate planning and utilization of the V2G system to support grid stability and efficiency by providing a clear and immediate measure of the amount of energy that can be drawn from or supplied to the grid through these parked EVs. This also implies that the flexibility resources provided by V2G systems are determined by the CS utilization.

The CS utilization with the decision time interval $\varDelta T$ (e.g., 1 h) is presented as [20]

$$\rho_t = \frac{\lambda_t}{m}\min\{h_t, \varDelta T\} + \sum_{i=-\infty}^{t-1}\frac{\lambda_i}{m}\min\{h_{t,i}^R, \varDelta T\}, \tag{3}$$

where m, λ_t, and λ_t indicate the number of chargers, the EV arrival rate in time slot t, and its service time. In Equation (3), the first term represents the CS utilization owing to new EVs arriving in timeslot t, while the second term accounts for the CS utilization resulting from existing EVs that arrived before timeslot t and continue to be serviced. The residual service time of the existing EVs $h_{t,i}^R$ is calculated as $h_{t,i}^R = \max\{h_i - (t-i)\varDelta T, 0\}$.

Considering the blocking probability, which indicates the likelihood of unaccommodated EVs owing to the full occupancy of chargers, the effective utilization of a CS equipped with m during timeslot t is adjusted.

$$\hat{\rho}_t = \rho_t\{1 - P_b(\rho_t, m)\}, \tag{4}$$

where the blocking probability is measured as $P_b(\rho_t, m) = {\rho_t^m/m!}\Big/{\sum_{i=0}^m \rho_t^i/i!}$ [21].

Using the effective utilization and charging power of the charger, c kW, the V2G flexibility resource at time slot t, C_t, is given by

$$C_t(\rho_t, \varDelta C) = c\varDelta Cm\rho_t\{1 - P_b(\rho_t, m)\}, \tag{5}$$

where $\varDelta C$ $(\leq \varDelta T)$ indicates the participation time duration in the V2G operation.

In Equation (5), the V2G flexibility resource represents the maximum capacity of the V2G system. Therefore, the V2G charging or discharging energy, denoted as q_t must be determined within the limits of the flexibility resource, as follows:

$$-C_t(\rho_t, \varDelta T) \leq a_t \leq C_t(\rho_t, \varDelta T) \ \ \forall t \in \mathcal{T}, \tag{6}$$

where $\mathcal{T}$ denotes the V2G operation time horizon; i.e., $\mathcal{T} = \{1, \cdots, t, \cdots, T\}$. Considering the charger efficiency $\eta \in (0,1]$, the actual charging or discharging quantity q_t is measured as

$$q_t \ = \begin{cases} a_t/\eta, & \text{if } a_t \geq 0, \\ \eta\, a_t, & \text{if } a_t < 0. \end{cases} \tag{7}$$

Moreover, unlike electrical energy storage systems, which may have a fixed capacity, the V2G flexibility resource changes with each time slot t based on the status of the chargers and the provision of V2G resources. If discharging is performed for the V2G, then the charging must return to the original level. Therefore, the service time increases by twice the participation time as $h_t^+ = h_t + 2\varDelta C$. This increases the CS utilization at time slot t and continues with the CS utilization in the next time slot.

To reflect these characteristics, a blocking probability constraint is considered as a V2G system constraint,

$$p_B(\rho_i^+, m) - P_b(\rho_i, m) \leq \zeta_{Th}, \quad i \in \{t, t+1, t+2, \cdots, T\}, \tag{8}$$

where ζ_{Th} refers to the blocking probability threshold. The blocking probability threshold is determined by the V2G service provider based on the system environment.

3.2. V2G Operation Problem

The goal of this study was to design a V2G operation method for managing SPG forecast errors. This is formulated as an SPG forecast error minimization problem. This study considers the mean squared error (MSE) as an error-management performance metric.

During the V2G operation time horizon, the MSE is calculated as

$$\mathcal{O}(\mathbf{a}) = \frac{1}{T}\sum_{t\in\mathcal{T}}(e_t - q_t)^2 = \frac{1}{T}\sum_{t\in\mathcal{T}}\epsilon_t^2, \tag{9}$$

where $\mathbf{a} = \{a_1, \cdots, a_t, \cdots, a_T\}$.

Considering the V2G operation constraints, the SPG forecast error minimization problem can be expressed as an optimization problem, as follows:

$$\begin{array}{cc} \max\limits_{\mathbf{a}} & \mathcal{O}(\mathbf{a}) \\ \text{subject to} & \text{Equations (6) and (8).} \end{array} \tag{10}$$

The problem presented in Equation (10) is a quadratic problem bounded by convex constraints. Consequently, if information such as the SPG and its forecast error for future time slots is known, then the problem can be addressed using search algorithms such as the gradient descent method in an iterative manner [22]. However, the premise of having foresight into future information violates the principle of causality and is infeasible in real-world scenarios [23]. This study introduces an effective approach to tackling the problems outlined in Equation (10) without the need for advanced knowledge of future time slots by employing DRL approaches.

4. Proposed V2G Operation Method

4.1. Markov Decision Process

As expressed in Equation (9), the V2G operation to manage the SPG forecast error is a sequential decision-making problem (SDP). The MDP is a general formalization of the SDP and a mathematical form of the RL problem [24]. To determine the optimal criterion for an MDP model, the state–action space and transition probability between spaces are required. However, the RL method eliminates the requirement for the transition probability. Therefore, only a definition of the state–action space is required to design the criteria for the RL method.

The state represents the environment required for V2G operation. For the object functions in Equation (9), the environmental variables used to determine the V2G operation action include the decision time t and the SPG forecast error e_t. Moreover, the V2G operation action is in the range of the flexibility resource that is based on the CS utilization ρ_t, as expressed in Equations (6) and (8), respectively. Therefore, the state space contains three tuples: the operation time, SPG forecast error, and CS utilization.

$$s_t = \{t, e_t, \rho_t\} \in \mathcal{S}. \tag{11}$$

The action space is defined as all the available action, considering CS utilization:

$$a_t \in \mathcal{A}. \tag{12}$$

The action space is a discrete space with size N:

$$\mathcal{A} = \{a^1, \cdots, a^j, \cdots, a^N\}. \tag{13}$$

It is noteworthy that the MDP-based RL method is solved only under discrete conditions. This implied that the state and action spaces are discrete spaces. However, a continuous state space can also be applied to the DRL method.

Moreover, the action at each stage should be determined within the state range, as follows:

$$s_{t+1} \leftarrow < s_t, a_t >. \tag{14}$$

Therefore, the feasible action set at time t is a subset of $\mathcal{A}$, considering the constraints in Equations (6) and (8):

$$\mathcal{A}_t \subseteq \mathcal{A}. \tag{15}$$

4.2. Proposed RL-based V2G Operation Method

An RL-based V2G operation is the decision-making process for the action at each stage s_t, considering the feasible action range $\mathcal{A}_t$.

The aim of V2G operation is to minimize the SPG forecast error, which is presented as the MSE during the operation time horizon, as shown in Equation (9). The objective function at each operation time t is modified as follows:

$$\begin{aligned} \mathcal{O}_t(a_t|e_t) &= \frac{1}{T}\sum_{i=1}^{T} \epsilon_t^2 \\ &= \frac{1}{T}\epsilon_t^2 + \frac{1}{T}\sum_{i=t+1}^{T} \hat{\epsilon}_i^2 \\ &= \frac{1}{T}\epsilon_t^2 + \mathcal{O}_{t+1}(\hat{a}_{t+1}|\hat{e}_{t+1}), \end{aligned} \tag{16}$$

where the values with a hat, $\hat{a}_{t+1}$ and $\hat{e}_{t+1}$, present the expected values.

From Equation (16), the rewards and returns of the RL method are defined. The reward at the operation time is the instantaneous value from the current action within the current state, and the return is the cumulative reward from the onward operation time. The reward at operation time t, r_t, is expressed as the first term in Equation (16):

$$r_t = \frac{1}{T}\epsilon_t^2. \tag{17}$$

Using the reward, the return R_t is defined:

$$\begin{aligned} R_t &= r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \cdots + \gamma^{T-t} r_T \\ &= r_t + \gamma R_{t+1}, \end{aligned} \tag{18}$$

where γ indicates the discount factor in $(0,1]$ that reduces the risk of the expected value from the decision time onward. The return in Equation (18) becomes the discounted objective function in Equation (16). Therefore, the decision making in the ESS operation action involves designing the action selection to minimize the reward, which is the error performance.

The state–action value function that represents the performance of a decided action in a given state is defined as follows:

$$\begin{aligned} Q(s_t, a_t) &= \mathbb{E}[R_t|s_t, a_t] \\ &= \mathbb{E}[r_t + \gamma Q(s_{t+1}, a_{t+1})|s_t, a_t]. \end{aligned} \tag{19}$$

If π represents the decision-making strategy of the action, then the optimal strategy is to minimize the state-action value of all states, $\pi^* = \operatorname{argmin}_\pi Q(s_t, a_t)$, $\forall\, s_t \in \mathcal{S}, a_t \in \mathcal{A}$. From the aspect of the state–action value function, it is expressed as $Q^*(s_t, a_t) = \min_\pi Q_\pi(s_t, a_t)$, $\forall\, s_t \in \mathcal{S}, a_t \in \mathcal{A}$. $Q_\pi(s_t, a_t)$ expresses the state–action value when the policy π is applied. From the Bellman optimality equation for the state–action function [25], the optimal state–action function can be rewritten as

$$\begin{aligned} Q^*(s_t, a_t) &= \mathbb{E}\left[r_t + \gamma \min_{a_{t+1}\in A_{t+1}} Q(s_{t+1}, a_{t+1}) | s_t, a_t\right] \\ &= \mathbb{E}[r_t + \gamma Q^*(s_{t+1}, a_{t+1}) | s_t, a_t]. \end{aligned} \tag{20}$$

The optimal state–action function in Equation (20) shows that the optimal policy is to determine the local optimal action at each decision time t because the expected reward from the onward decision time is the optimal value. Therefore, the optimal action is determined as follows:

$$\mathrm{a}_\mathrm{t}^* = \underset{a_t\in\mathcal{A}_t}{\operatorname{argmin}}\, Q^*(s_t, a_t). \tag{21}$$

If the state–action probability at each decision time is known, then the optimal state–action function in Equation (20) is calculated, and the optimal action in Equation (21) is determined using the optimal state–action function. This requires information from all the states, including future times, making the method impractical. In this study, the state–action function is estimated through learning.

In tabular-based RL methods such as Q-learning, the state–action value functions are updated as

$$Q(s_t, a_t) \leftarrow \quad (1-\alpha)Q(s_t, a_t) + \alpha\left[r_t + \gamma \min_{a\in\mathcal{A}_{t+1}} Q(s_{t+1}, a)\right], \tag{22}$$

where α is a learning rate of convergence in $(0,1]$. This is a linear approximation function.

In the DRL approach, the state–action value function is approximated using deep-learning-based Q-network with parameter θ_t^k, which is a nonlinear approximation function [26]. The Q-network is updated to minimize the quadratic loss function. The quadratic loss function of the k-th iteration at decision time t is expressed as follows:

$$L_t^k(\theta_t^k) = \frac{1}{2}\{y_t - Q(s_t, a_t; \theta_t^k)\}^2, \tag{23}$$

where $y_t = \mathbb{E}\left[r_t + \gamma \min_{a_{t+1}\in A_{t+1}} Q(s_{t+1}, a_{t+1}; \theta_t^{k-1}) | s_t, a_t\right]$. The Q-network parameter θ_t^k is updated based on the gradient descent method, as follows:

$$\begin{aligned} \theta_t^k &= \theta_t^{k-1} - \alpha \nabla_{\theta_t^k} L_t^k(\theta_t^k) \\ &= \theta_t^{k-1} - \alpha\{y_t - Q(s_t, a_t; \theta_t^k)\}\nabla_{\theta_t^k} Q(s_t, a_t; \theta_t^k). \end{aligned} \tag{24}$$

When the state–action value function is iterated to convergence, the action policy is determined using the approximate value:

$$\mathrm{a}_\mathrm{t} = \underset{a_t\in\mathcal{A}_t}{\operatorname{argmin}}\, \mathrm{Q}(s_t, a_t; \theta_t^k). \tag{25}$$

Three engineering techniques were applied to enhance the DRL method. The first is a penalty-based reward system. To determine the actions in Equation (25), it is necessary to compute the feasible action space $\mathcal{A}_t$. Given that this feasible space is present in recursive form, as shown in Equation (8), repeated calculations are required. To streamline this process, the action space to be decided at each decision time is relaxed to $\mathcal{A}$ and the feasibility of the determined actions is subsequently assessed. Based on this framework, the reward is modified as

$$r_t \quad = \begin{cases} r_t, & \text{if a feasible } a_t, \\ \kappa, & \text{otherwise,} \end{cases} \tag{26}$$

where κ is the penalty value. This approach simplifies the decision-making process by broadening the initial set of possible actions and applying criteria to evaluate their feasibility, thereby reducing the computational complexity associated with the identification of viable actions within a recursive framework.

The second engineering skill involves the use of the replay buffer $\mathcal{D}$. SPG exhibits non-stationary characteristics owing to weather changes, which can significantly hinder the convergence of DRL models. A replay buffer is used to address this challenge. This buffer stores previous states, actions, and rewards, facilitating the learning of the state–action value function approximation through a random sampling of these stored experiences. The use of random sampling in the replay buffer helps mitigate the correlation between sequential data points, thereby enhancing the efficiency of DRL learning by providing a more diversified and representative sample of experiences for training. The technique is crucial for adapting DRL methods to environments with non-stationary dynamics, such as those influenced by variable weather conditions affecting SPG.

Finally, a target network with a delayed update approach is implemented. The loss function for the state–action value function approximation in Equation (23) is composed of the target value y_t and the action value $Q(s_t, a_t; \theta_t^k)$, with updates proceeding sequentially based on the gradient descent method. The target network with parameter Θ_t^k is a clone of the main state–action value function approximation network with parameter θ_t^k. It is used to calculate the target values, as follows:

$$y_t = \mathbb{E}\left[r_t + \gamma \min_{a_{t+1} \in A_{t+1}} Q(s_{t+1}, a_{t+1}; \Theta_t^{k-1}) | s_t, a_t\right], \tag{27}$$

The target network is updated less frequently than the main network. This separation creates a more stable learning target, reducing the variance in updates and mitigating the risk of divergent behavior that can occur in a rapidly changing environment.

The proposed DRL-based V2G operation method is summarized as follows:

A DRL-based V2G operation method

```
Initialization
    Initialize the main state–action value function network parameters $\theta$.
    Copy the parameter to the target network $\Theta = \theta$.
    Initialize replay buffer $\mathcal{D}$, the maximum number of iterations $K$, the target
network update frequency $Z$

Optimal policy learning
    For $k = 1$ to $K$ do
        Set the current state to $s_0$.
        For $t = 1$ to $T$ do
            Replay buffering
            Decide the action using Equation (25) with exploration process.
            Calculate the reward using Equation (26).
            Store $(s_t, a_t, r_t, s_{t+1})$ as a set of sample data in $\mathcal{D}$.
            If $r_t == \kappa$ then
                Infeasible terminated.

            Policy learning
            Sample random minibatch of transitions $(s_t, a_t, r_t, s_{t+1})$ from $\mathcal{D}$.
            Calculate the target value $y_t$ using Equation (27).
            Update a main network parameter $\theta_t^k$ using Equation (24).
        end for

        Target network updating
        If $k$ mode $Z$ then
            Copy the main network parameter $\theta$ to the target network $\Theta = \theta$.
    end for
```

The architecture of the proposed DRL-based V2G operation method is expressed in Figure 2.

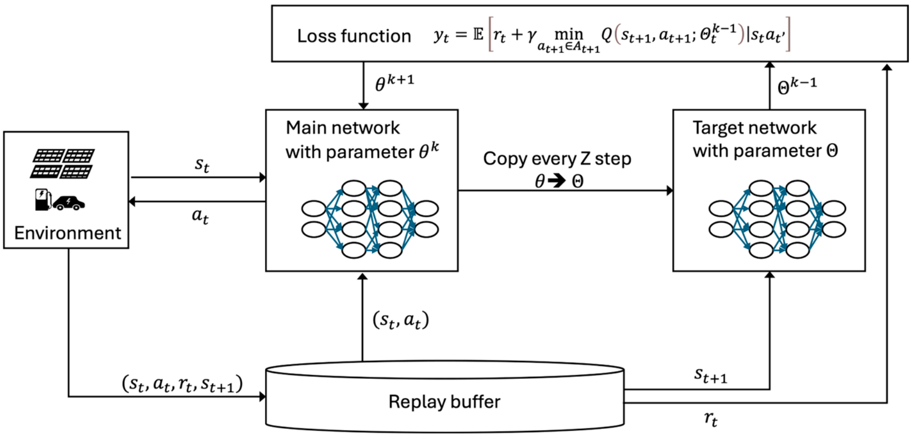

Loss function $y_t = \mathbb{E}\left[r_t + \gamma \min_{a_{t+1} \in A_{t+1}} Q\left(s_{t+1}, a_{t+1}; \Theta_t^{k-1}\right) | s_t a_{t'}\right]$
θ^{k+1}
Θ^{k-1}
Main network with parameter θ^k
Copy every Z step
$\theta \rightarrow \Theta$
Target network with parameter Θ
s_t
a_t
Environment
(s_t, a_t)
(s_t, a_t, r_t, s_{t+1})
Replay buffer
s_{t+1}
r_t

Figure 2. The architecture of the proposed DRL-based V2G operation method.

Our research employs a model-free approach based on DRL, which optimizes actions within a given state–action space through deep learning techniques. The objective of DRL is to discover the optimal action sequence that minimizes a cumulative reward by learning from interactions with the environment, without requiring a model of the environment's dynamics.

The state space in our study is defined by two critical aspects: the instantaneous error in SPG forecasts and the utilization of CSs. The instantaneous SPG error provides real-time deviation information between forecasted and actual power generation, which is crucial for adjusting control strategies dynamically. The CS utilization is modeled using an M/M/C queue-based Markov chain model. This modeling approach allows us to capture the operational dynamics of multiple CSs under varying load conditions, providing a realistic framework for simulating V2G interactions.

Each decision point in our system adheres to the characteristics of an MDP, where the choice made at any given moment is influenced solely by the current state and not by any previous states. This property ensures that our DRL algorithm can efficiently navigate the decision-making process, optimizing actions based on current observations without the burden of historical data.

The model-free nature of the DRL algorithm implies that it does not require a predefined model of the environment, making it highly adaptable to various operational contexts. By adjusting parameters within the DRL framework, the algorithm can be tailored to different environmental conditions and constraints, enhancing its applicability across different V2G scenarios.

This methodological foundation supports our V2G operation strategy, allowing for the sophisticated and dynamic management of energy resources in real time. By leveraging the capabilities of DRL, our approach addresses the inherent uncertainties and variabilities in renewable energy systems, offering a robust tool for enhancing grid reliability and efficiency.

5. Results and Discussion

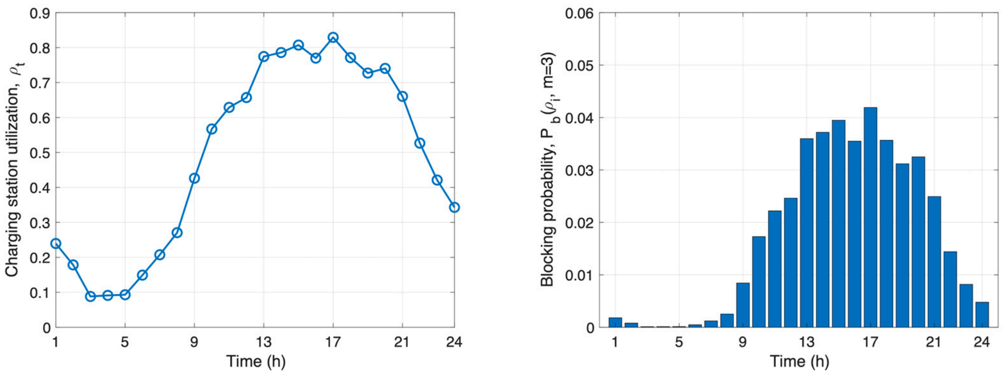
5.1. Experimental Environments

The efficacy of our study was assessed using SPG data with a capacity of 187 kW alongside forecasting data recorded by the Korea Southern Power Corporation from 2013 to 2022. The data were made available through a public data portal operated by the Ministry of the Interior and Safety, Korea, and formed the empirical basis of our research [27]. Data spanning from 2013 to 2020 were employed to train the proposed DRL-based V2G operation strategy, while data from 2021 to 2022 were utilized to evaluate the effectiveness of the model.

To enhance the clarity of our analysis and facilitate a more accessible interpretation of the computational results, we normalized the SPG capacity to 100 kWh. This normalization, a standard procedure in modeling and simulation efforts, aims to standardize the data scale, thereby simplifying the comparative analysis and evaluation processes. Notably, the MSE for the SPG forecast in the absence of V2G operations stands at 44.38. This metric serves as a critical benchmark against which the performance improvements introduced by the proposed DRL-based V2G strategy are measured, underscoring the potential of our approach in enhancing the accuracy and reliability of SPG forecasting.

For our V2G system analysis, we utilized CS information sourced from the Korea Environment Corporation [20], which offers critical insights into the operational dynamics of the CS infrastructure. The selected CS was equipped with three DC chargers, each boasting a charging power of 50 kWh, which served as a practical foundation for evaluating the proposed V2G operation strategy.

As illustrated in Figure 3, the CS status provides a comprehensive view of the operational capacity and limitations of the charging infrastructure. Notably, Figure 3a shows that the maximum utilization rate of the CS approaches approximately 0.83, with an average utilization per charger of 0.28. This utilization metric is pivotal because it indicates the extent to which CS resources are leveraged under typical operational conditions. Figure 3b shows the maximum blocking probability associated with CS, recorded at 0.04. This figure signifies a 4% likelihood that an arriving EV will encounter a fully occupied CS, thereby being unable to initiate immediate charging.



(**a**) (**b**)

Figure 3. Charging station status. (**a**) Charging station utilization; (**b**) blocking probability of the charging station.

In our study, the approximation of the state–action value function, which is crucial to the DRL framework, was facilitated by a fully connected neural network (FNN) architecture comprising two hidden layers. Each of these layers was densely populated with 256 neurons, providing the necessary computational capacity to effectively capture the complex dynamics of the V2G operation strategy.

For the iterative optimization of our neural network, the training process employed a minibatch size of 60 for each gradient update. This specific batch size was chosen to synchronize updates on a bi-monthly basis, aligning with the operational timelines typical of V2G system evaluations. Furthermore, a cycle for updating the target network was established over six iterations. When combined with the bi-monthly update schedule and the chosen minibatch size, this configuration implies that the target network underwent an update approximately once per year. This temporal setting was strategically selected to mimic real-world operational and decision-making cycles within the context of V2G systems.

The learning algorithm was fine-tuned with a learning rate of 0.001, ensuring a gradual and stable convergence toward optimal policy solutions. These parameters were carefully selected to optimize the performance and reliability of the proposed DRL-based framework with the aim of delivering robust and practical solutions for managing the intricate interplay between SPG forecasts, EV charging/discharging activities, and CS utilization within the V2G ecosystem.

We conducted our experiments on a system equipped with a single NVIDIA GeForce RTX A6000 GPU. Moreover, we utilized the PyTorch framework due to its robust support for tensor operations and compatibility with GPU architectures. To ensure the reliability and reproducibility of our results, all experimental outcomes were obtained by averaging the results of 200 repeated trials. This approach mitigated any anomalies or outliers that could have affected the stability and accuracy of our findings.

Table 1 provides a detailed description of the implementation parameters used in our study, outlining their specific values and roles to ensure clarity and reproducibility in our experimental setup.

Table 1. Implementation parameters.

Name	Values
System parameters	
Number of charging stations	3, 4, 5
Blocking probability threshold	0.02, 0.04, 0.06, 0.08, 0.10
Model parameters	
Network model	FNN
Hidden layer	2
Neurons	256, 256
Learning rate	0.001
Discount factor	0.01, 0.20, 0.40, 0.60, 0.80, 0.99
Minibatch size	60
Target network update cycle	6

5.2. MSE Reduction Performance

Table 2 presents a comprehensive analysis of the MSE changes in the SPG forecasts under the application of the proposed V2G operation strategy across varying blocking probability thresholds and discount factors. Additionally, Table 3 presents the MSE reduction ratios relative to the baseline scenario in which V2G operations are not utilized.

Table 2. MSE change in SPG forecast applying the proposed V2G operation method.

Blocking probability threshold, ζ_{Th}	**Discount factor, γ**					
	0.01	**0.20**	**0.40**	**0.60**	**0.80**	**0.99**
0.02	24.44	24.01	23.99	24.05	24.40	26.14
0.04	24.18	24.18	22.95	23.63	23.79	25.24
0.06	22.77	21.56	19.93	19.95	20.98	22.64
0.08	23.81	20.66	20.30	19.65	19.45	20.79

Table 3. MSE reduction ratio applying the proposed V2G operation method.

Blocking probability threshold, ζ_{Th}	**Discount factor, γ**					
	0.01	**0.20**	**0.40**	**0.60**	**0.80**	**0.99**
0.02	44.9%	45.9%	45.9%	45.8%	45.0%	41.1%
0.04	45.5%	45.5%	48.3%	46.8%	46.4%	43.1%
0.06	48.7%	51.4%	55.1%	55.0%	52.7%	49.0%
0.08	46.3%	53.4%	54.2%	55.7%	56.2%	53.1%

By implementing the V2G operation strategy, we recorded the SPG forecast errors falling between 19.45 and 26.14. This signifies an improvement over the baseline scenario, which had an MSE of 44.38, demonstrating an error reduction performance ranging from over 40% to as much as 56% across all cases.

5.3. Effect of Characteristics

5.3.1. Effect of Blocking Probability Threshold

A notable trend observed in the results in Tables 2 and 3 is a reduction in MSE as the blocking probability threshold increases, indicating an enhanced accuracy in SPG forecasting owing to the application of the V2G operation method. This phenomenon can be attributed to the interplay between the imposed constraints and resultant flexibility within the V2G system. The blocking probability threshold serves as a critical constraint that directly affects CS utilization. This effectively delineates the operational limits within which the CS can accommodate EV charging and discharging activities without compromising service availability. A higher blocking probability threshold signifies a greater tolerance for reaching or exceeding the capacity of the CS, thereby allowing for increased flexibility in V2G operations. This flexibility is crucial for dynamically managing the integration of EVs into the grid, facilitating optimal charging and discharging schedules that align with the SPG forecast errors and grid demands. Moreover, from the perspective of CSs, an increased blocking probability threshold may complicate EV charging, potentially degrading the quality of charging services. Therefore, V2G service providers must consider this trade-off when setting the blocking probability thresholds.

5.3.2. Effect of Discount Factor

In Table 2, the MSE of the SPG forecast varies based on the discount factor in addition to the blocking probability threshold. To examine the impact of the discount factor on the MSE of the SPG forecasts more closely, Figure 4 aggregates the results across different blocking probabilities to depict the variation in MSE values that correspond to each discount factor. Figure 4 shows that the MSE of the SPG forecasts demonstrates the lowest value within the discount factor range of 0.4 to 0.6.

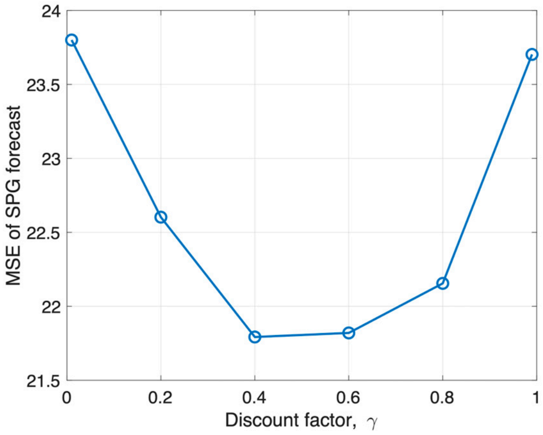


Figure 4. Average MSE of SPG forecast applying the proposed V2G operation method, varying the discount factor.

The discount factor, which is a critical parameter in DRL algorithms, essentially determines the weight assigned to future rewards compared with immediate rewards. A higher discount factor values future rewards more, encouraging strategies that may benefit long-term outcomes, even if they do not yield immediate gains. Conversely, a lower discount factor prioritizes immediate rewards, which can be beneficial in environments in which quick adaptation is essential for performance. In conventional DRL-based resource management systems, particularly those involving BESSs, the absence of uncertainty in resource availability often justifies setting a high discount factor, such as 0.9–1 [16–18].

This approach favors long-term benefits and enhances performance by significantly valuing future rewards over immediate ones. The rationale is that, in scenarios with predictable resource behavior, prioritizing future outcomes can effectively optimize system operations and energy utilization over extended periods.

However, the V2G system introduces a different dynamic, given the inherent flexibility and variability associated with EV behavior and availability. V2G systems offer dynamic flexibility resources influenced by numerous unpredictable factors, such as EV owners' charging habits, mobility patterns, and willingness to participate in V2G services. These elements introduce a level of uncertainty that differs markedly from the more stable characteristics of BESSs. A discount factor within 0.4–0.6 strikes a balance between the importance of immediate and future rewards, aligning with the operational realities of V2G systems. It ensures that the DRL model is adequately sensitive to adapt to the short-term fluctuations and opportunities presented by the availability of EVs for charging or discharging without overly discounting the significance of forthcoming rewards. This balanced approach facilitates the effective management of SPG forecast errors by leveraging the unique characteristics of V2G systems to enhance grid stability and renewable energy integration. This adjustment underscores the necessity of considering future values amidst resource uncertainty, offering crucial insights into managing systems with DRL algorithms. Identifying the optimal discount factor that accounts for these uncertainties represents a significant area for future research. Investigating this will not only improve the robustness of DRL applications in variable and uncertain environments but also enhance the overall efficiency and effectiveness of renewable energy integration strategies.

5.3.3. Combining Effects of Blocking Probability Threshold and Discount Factor

To examine the combined effects of the blocking probability threshold and discount factor on the V2G operations, Figure 5 presents the distribution of the maximum blocking probability increments in the form of boxplots. These plots delineate the 25% and 75% quantile values as the lower and upper bounds of the box, respectively, and the median (50% quantile) is represented by the central line. Figures 5a,b specifically illustrate scenarios with discount factors set at 0.4 and 0.99, respectively.

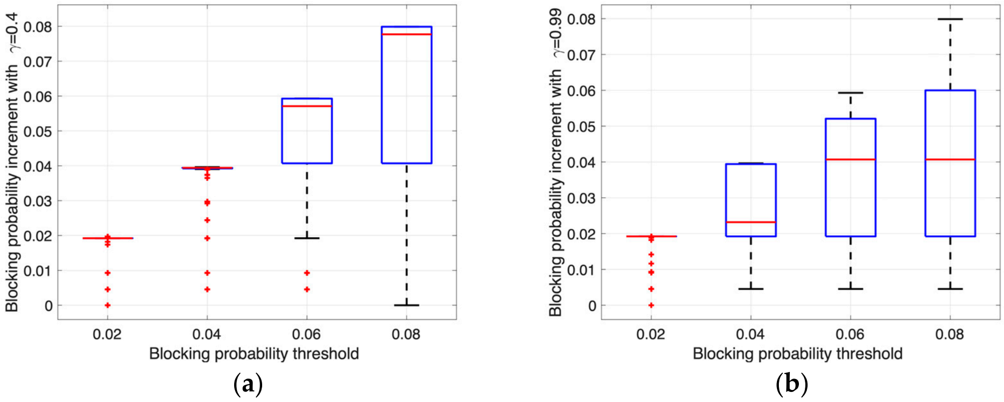


Figure 5. Maximum blocking probability increment distribution: (**a**) 0.4 discount factor case; (**b**) 0.99 discount factor case.

Figure 5a shows that with a discount factor of 0.4, the average value of the maximum blocking probability increment increases closer to the threshold value as the blocking probability threshold itself increases. This trend indicates that lower discount factors that prioritize immediate rewards allow for increased system flexibility without significantly compromising future outcomes. However, for a discount factor of 0.99, as shown in Figure 5b, even as the blocking probability threshold increases, the average value of the maximum blocking probability increment tends to converge. This phenomenon suggests that a high discount factor, which heavily weighs future values, may not capitalize on the immediate benefits of increased flexibility resulting from higher blocking probability thresholds.

A closer inspection of the minimum values of the blocking probability increment reveals that at a blocking probability threshold of 0.8, scenarios with a discount factor of 0.99 exhibit larger values compared to those with a discount factor of 0.4. A lower blocking probability increment implies a less effective V2G operation, suggesting that a higher discount factor may not always facilitate optimal V2G interactions. Consequently, these results can be interpreted to mean that a discount factor of 0.99, compared with 0.4, may offer more stable operation within V2G systems, despite the potentially less effective utilization of V2G flexibility.

5.4. Discussion and Summary

This section presents a thorough analysis of the proposed V2G operation strategy, focusing on its impact on SPG forecast accuracy, with special attention paid to the roles of the blocking probability thresholds and discount factors. The key aspects can be summarized as follows:

First, the study demonstrates a significant reduction in the MSE of the SPG forecasts when the V2G operation method is applied across various blocking probability thresholds and discount factors. The MSE improvement, ranging from over 40% to as much as 56%, compared to a baseline scenario without V2G utilization, highlights the effectiveness of the proposed strategy in enhancing the SPG forecast accuracy. This improvement underscores the potential of V2G systems in mitigating the variability and uncertainty inherent in solar power generation, thereby contributing to more stable and efficient grid management.

Second, the analysis reveals the critical influence of the blocking probability threshold on the efficacy of V2G operation. An increase in the threshold enables for greater system flexibility by accommodating more EV charging and discharging activities, which, in turn, leads to enhanced SPG forecast accuracy. This finding suggests that strategically managing the blocking probability threshold can optimize V2G operations, balancing the need for CS utilization with the benefits of increased renewable energy integration into the grid.

Finally, this study examined the combined effects of blocking probability thresholds and discount factors on V2G operations. It is observed that lower discount factors, which prioritize immediate rewards, facilitate an increase in system flexibility and consequently improve SPG forecast accuracy. However, a higher discount factor, while potentially offering a more stable V2G system operation, may not capitalize on the immediate benefits of the increased flexibility afforded by higher blocking probability thresholds. This interplay between the discount factor settings and blocking probability thresholds highlights the importance of carefully calibrating these parameters to achieve optimal performance in V2G systems.

In summary, the discussion emphasizes the potential of a well-designed V2G operation strategy that utilizes DRL techniques to significantly improve SPG forecast accuracy. Moreover, it highlights the necessity of balancing operational flexibility with strategic foresight through adept manipulation of key system parameters: namely, blocking probability thresholds and discount factors. This balance is crucial for maximizing the benefits of V2G systems in supporting renewable energy integration and enhancing grid stability.

6. Conclusions

This study proposed a DRL-based V2G operation strategy for managing SPG forecast errors. By intricately considering the status of the CS, the developed DRL-based strategy significantly enhanced the accuracy of SPG forecasts, demonstrating a potential reduction in MSE of up to 56% compared with the scenario without V2G. This improvement not only highlights the effectiveness of the proposed strategy in optimizing renewable energy use but also underlines the critical role of V2G systems in stabilizing the grid amidst the variability inherent in solar power. This study further discusses the dynamics between blocking probability thresholds and discount factors within the V2G operation framework. The findings reveal an optimal balance that maximizes system flexibility and forecast accuracy, suggesting a strategic pivot toward moderate discount factor values to accommodate the unpredictable nature of EV availability and behavior. This balance is crucial to leverage EVs as dynamic energy resources, thereby enabling more resilient and efficient grid operations.

Based on the foundations of this study, there are several promising directions for future research. First, by extending the findings that emphasize the importance of tailoring discount factors based on resource uncertainty, future studies could explore the development of DRL methodologies that incorporate discount factors. Such an approach would dynamically adjust the discount factor in response to the fluctuating uncertainty levels of renewable energy resources, potentially enhancing the adaptability and effectiveness of DRL strategies for managing renewable energy integration. Furthermore, while this research primarily focuses on a simplified model for the CS status, there is a significant opportunity to delve into more complex models of CS operation. Future work could investigate the application of V2G strategies within the context of an equivalent circuit model, offering a more comprehensive understanding of CS dynamics. This advanced model could provide new insights into optimizing V2G operations, particularly in scenarios characterized by complex interactions between EVs, renewable energy sources, and grid infrastructure.

Additionally, we encountered challenges in applying advanced DRL models like DDPG or twin-delayed DDPG in the V2G context due to their complex network structures and high sensitivity to parameter tuning. The inherent uncertainties in SPG and the dynamic demands of V2G systems often led to convergence issues. Future research should focus on enhancing the robustness of these models to ensure reliable convergence in the face of such uncertainties. Developing methodologies that can adaptively manage the complexities of real-world energy systems, while maintaining stability and performance, will be crucial. Addressing these challenges will advance the application of DRL in energy management, contributing to more reliable and efficient renewable energy integration in the power grid.

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

Funding: This research was supported by the Korea Electric Power Corporation (Grant Number: R22XO02-23).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found at https://www.data.go.kr/data/15043386/fileData.do (accessed on 11 April 2024).

Conflicts of Interest: Author Moon-Jong Jang was employed by the company Korea Electric Power Corporation. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest. The authors also declare that this study received funding from Korea Electric Power Corporation. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

References

1. International Energy Agency (IEA). Renewables 2023—Analysis and forecasts to 2028. 2024. Available online: https://www.iea.org/reports/renewables-2023/ (accessed on 11 April 2024).
2. Kabeyi, M.J.B.; Olanrewaju, O.A. The levelized cost of energy and modifications for use in electricity generation planning. *Energy Rep.* **2023**, *9*, 495–534.
3. Nie, Y.; Li, X.; Paletta, Q.; Aragon, M.; Scott, A.; Brandt, A. Open-source sky image datasets for solar forecasting with deep learning: A comprehensive survey. *Renew. Sustain. Energy Rev.* **2024**, *189*, 113977.
4. Krishnan, N.; Kumar, K.R.; Inda, C.S. How solar radiation forecasting impacts the utilization of solar energy: A critical review. *J. Clean. Prod.* **2023**, *388*, 135860.
5. International Energy Agency (IEA). Global EV Outlook 2023—Catching up with Climate Ambitions. 2024. Available online: https://www.iea.org/reports/global-ev-outlook-2023/ (accessed on 11 April 2024).
6. Alanazi, F. Electric vehicles: Benefits, challenges, and potential solutions for widespread adaptation. *Appl. Sci.* **2023**, *13*, 6016.
7. Panchanathan, S.; Vishnuram, P.; Rajamanickam, N.; Bajaj, M.; Blazek, V.; Prokop, L.; Misak, S. A comprehensive review of the bidirectional converter topologies for the vehicle-to-grid system. *Energies* **2023**, *16*, 2503.
8. Ferris, M.C.; Philpott, A. Renewable electricity capacity planning with uncertainty at multiple scales. *Comput. Manag. Sci.* **2023**, *20*, 41.
9. Gheouany, S.; Ouadi, H.; Giri, F.; El Bakali, S. Experimental validation of multi-stage optimal energy management for a smart microgrid system under forecasting uncertainties. *Energy Convers. Manag.* **2023**, *291*, 117309.
10. Srinivasan, A.; Wu, R.; Heer, P.; Sansavini, G. Impact of forecast uncertainty and electricity markets on the flexibility provision and economic performance of highly-decarbonized multi-energy systems. *Appl. Energy* **2023**, *338*, 120825.
11. Song, B.; Jin, W.; Li, C.; Khakichi, A. Economic management and planning based on a probabilistic model in a multi-energy market in the presence of renewable energy sources with a demand-side management program. *Energy* **2023**, *269*, 126549.
12. Zheng, X.; Bai, F.; Zhuang, Z.; Chen, Z.; Jin, T. A new demand response management strategy considering renewable energy prediction and filtering technology. *Renew. Energy* **2023**, *211*, 656–668.
13. Li, B.; Liu, Z.; Wu, Y.; Wang, P.; Liu, R.; Zhang, L. Review on photovoltaic with battery energy storage system for power supply to buildings: Challenges and opportunities. *J. Energy Storage* **2023**, *61*, 106763.
14. Wongdet, P.; Boonraksa, T.; Boonraksa, P.; Pinthurat, W.; Marungsri, B.; Hredzak, B. Optimal capacity and cost analysis of battery energy storage system in standalone microgrid considering battery lifetime. *Batteries* **2023**, *9*, 76.
15. Alfaverh, F.; Denai, M.; Sun, Y. Optimal vehicle-to-grid control for supplementary frequency regulation using deep reinforcement learning. *Electr. Power Syst. Res.* **2023**, *214*, 108949.
16. Maeng, J.; Min, D.; Kang, Y. Intelligent charging and discharging of electric vehicles in a vehicle-to-grid system using a reinforcement learning-based approach. *Sustain. Energy Grids Netw.* **2023**, *36*, 101224.
17. Dong, J.; Yassine, A.; Armitage, A.; Hossain, M.S. Multi-Agent Reinforcement Learning for Intelligent V2G Integration in Future Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 15974–15983.
18. Wang, R.; Chen, Z.; Xing, Q.; Zhang, Z.; Zhang, T. A modified rainbow-based deep reinforcement learning method for optimal 598 scheduling of charging station. *Sustainability* **2022**, *14*, 1884.
19. Shibl, M.M.; Ismail, L.S.; Massoud, A.M. Electric vehicles charging management using deep reinforcement learning considering vehicle-to-grid operation and battery degradation. *Energy Rep.* **2023**, *10*, 494–509.
20. Jang, M.J.; Kim, T.; Oh, E. Data-Driven Modeling of Vehicle-to-Grid Flexibility in Korea. *Sustainability* **2023**, *15*, 7938.
21. Leon-Garcia, A. *Probability, Statistics, and Random Processes for Electrical Engineering*, 3rd ed.; Pearson: London, UK, 2021.
22. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004.
23. Oppenheim, A.V.; Willsky, A.S.; Nawab, S.H. *Signals and Systems*; Prentice Hall: Upper Saddle River, NJ, USA, 1997.
24. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018.
25. Kirk, D.E. *Optimal Control Theory: An Introduction*; Prentice Hall: Upper Saddle River, NJ, USA, 1970.
26. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533.
27. Ministry of the Interior and Safety, Korea. Public Data Portal. 2024. Available online: https://www.data.go.kr/index.do (accessed on 11 April 2024).

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.