Deep Reinforcement Learning-Based Allocation of Mobile Wind Turbines for Enhancing Resilience in Power Distribution Systems

Ruotan Zhang, *Student Member, IEEE,* Jinshun Su, *Student Member, IEEE,* Payman Dehghanian, *Senior Member, IEEE,* Mohannad Alhazmi, *Member, IEEE,* and Xiaoyuan Fan, *Senior Member, IEEE*

***Abstract*—The growing adoption of wind energy resources has demonstrated notable benefits in combating climate change. Mobile wind turbines (MWTs) are uniquely positioned to navigate transportation systems, being towed by trucks, and supply energy to power distribution systems (PDSs). This flexibility enables MWTs to serve as emergency power sources, thereby contributing to enhancing the system resilience by facilitating service restoration following extreme events. This paper presents a novel framework based on Multi-agent Deep Reinforcement Learning (MADRL) to dispatch MWTs for service restoration. Deep Q-learning (DQL) and Double Deep Q-learning (DDQL) approaches are implemented within the agent for training and comparison purposes. Additionally, an action limitation is incorporated into the proposed framework in order to mitigate the influence of wind power fluctuations. Case studies conducted on an integrated power-transport system, comprising a Sioux Falls transportation system and four IEEE 33-bus test systems, illustrate the effectiveness of the proposed restoration scheduling policy in enhancing PDSs' resilience against disasters.**

***Index Terms*—Mobile Wind Turbine (MWT), Double Deep Q-learning (DDQL), Deep Q-learning (DQL), Multi-agent Deep Reinforcement Learning (MADRL), resilience, service restoration.**

Nomenclature

Acronyms

AC	Actor critic
$AC-OPF$	Alternating current optimal power flow.
AI	Artificial intelligence.
AL	Action limitation.
AR	Agent reward.
CNN	Convolutional neural network.
DQL	Deep Q-learning.
DRL	Deep reinforcement learning.
$DDPG$	Deep deterministic policy gradient.
$DDQL$	Double deep Q-learning.
$DRLBRE$	DRL-based resilience enhancement.
DNN	Deep neural network.
GFM	Grid forming.
$HILP$	High-impact, low-probability.
HSS	Hydrogen storage system.
$H2P$	Hydrogen-to-power.
$IPPO$	Independent proxy policy optimization.
MDP	Markov decision process.
MPS	Mobile power source.
MWT	Mobile wind turbine.
$MADRL$	Multi-agent deep reinforcement learning.
$MILP$	Mixed integer linear programming.
PG	Policy gradient.
$P2H$	Power-to-hydrogen.
PDS	Power distribution system.
RC	Repair crew.
SR	System reward.
TS	Transportation system.
TN	Transportation node.

Indices and Sets

$i \in B$	Index and set of PDS buses.
$(i,j) \in L$	Index and set of PDS lines.
$d \in B_{ld}$	Index and set of loads.
$t \in T$	Index and set of time steps.

Parameters

c_d^{lo}	Load outage cost of demand at bus d ($\$/kWh$).
$\bar{P}_{d,t}^{ld}$	Baseline real power demand at bus d at time t (kW).
$\bar{Q}_{d,t}^{ld}$	Baseline reactive power demand at bus d at time t ($kvar$).
$\underline{V}$	Minimum permissible voltage ($p.u.$).
$\bar{V}$	Maximum permissible voltage ($p.u.$).
r_{ij}	Resistance of line (i,j) ($p.u.$).
x_{ij}	Reactance of line (i,j) ($p.u.$).
$\bar{S}_{ij}$	Capacity limit of line (i,j) (kVA).
μ_i^{h2p}	H2P conversion factor of HSS unit i.
η_i^{h2p}	H2P efficiency of HSS unit i.
μ_i^{p2h}	P2H conversion factor of HSS unit i.
η_i^{p2h}	P2H efficiency of HSS unit i.

Variables

$P_{d,t}^{ld}$	Real power demand at bus d at time t (kW).
$Q_{d,t}^{ld}$	Reactive power demand at bus d at time t ($kVAR$).

This work was supported in part by the NSF under Grant ECCS-2114100 and Grant RISE-2220626. The presented work was in support of NSF INTERN program, as part of 5G Energy FRAME project funded by the US Department of Energy Office of Science Advanced Scientific Computing Research (ASCR) Program. The work of M. Alhazmi was supported by the Ongoing Research Funding Program under project ORF-2025-635, King Saud University, Riyadh, Saudi Arabia.

R. Zhang, J. Su, P. Dehghanian are with the Department of Electrical and Computer Engineering, The George Washington University, Washington, DC 20052, USA (e-mails: zhangruotan114@gwu.edu; jsu66@gwu.edu; payman@gwu.edu).

M. Alhazmi is with the Department of Electrical Engineering, College of Applied Engineering, King Saud University, Riyadh 11421, Saudi Arabia (e-mail: mohalhazmi@ksu.edu.sa).

X. Fan is with the Eaton Research Labs, Golden, CO, USA. (e-mail: XiaoyuanFan@eaton.com).

$P_{g,t}^{mg}$	Real power supply from the main grid (kW).
$Q_{g,t}^{mg}$	Reactive power supply from the main grid ($kVAR$).
$P_{g,t}^{wt}$	Real power output of MWT (kW).
$Q_{g,t}^{wt}$	Reactive power output of MWT ($kVAR$).
$P_{ij,t}$	Real power flow of line (i,j) at time t (kW).
$Q_{ij,t}$	Reactive power flow of line (i,j) at time t ($kVAR$).
$V_{i,t}$	From bus voltage for line (i,j) ($p.u.$).
$V_{j,t}$	To bus voltage for line (i,j) ($p.u.$).
$e_{i,t}$	Binary variable indicating the energized status of bus i at time t (1 if energized, 0 otherwise).
$y_{ij,t}$	Binary variable indicating the energized status of line (i,j) at time t (1 if energized, 0 otherwise).
P_{i}^{h2p}	Power output of HSS unit i (kW).
E_{i}^{H}	Energy storage value at time step i (kWh).
E_{i-1}^{H}	Energy storage value at time step $i-1$ (kWh).
P_{i}^{p2h}	Hydrogen power input to HSS unit i (kW).
p_{i}^{p2h}	Supply power from PDS to HSS unit i (kW).

I. Introduction

CLIMATE change has led to an increased frequency and magnitude of HILP events, which have been witnessed to result in extensive equipment damage, prolonged electricity outages, and significant disruptions in modern society. Climate change-induced power outages have caused substantial economic losses and posed significant threats to human life, highlighting the urgent need to enhance the resilience of power grids to such extremes [1]. For example, Hurricane Maria in 2017 severely impacted Puerto Rico by disrupting 31 major power-generating units across 20 facilities, leaving the entire island without electricity [2]. Similarly, in February 2021, an extreme winter storm led to a widespread electricity generation failure in Texas, resulting in over 4.5 million households experiencing prolonged power outages and approximately $130 billion in economic losses [3]. Given these significant disruptions, a resilient electric system should prioritize the restoration of essential services, such as medical facilities and police stations [4]. Given the vulnerability of rural infrastructure to HILP events such as natural disasters-—which can damage TS, disrupt power distribution, and cause shortages of fossil fuels (e.g., gasoline, natural gas)—-renewable energy resources, which do not rely on fossil fuels, should be integrated into the restoration efforts to mitigate the impact of these events [5]–[7].

A. Literature Review

Due to their spatiotemporal flexibility, MPSs have become essential in enhancing the resilience in PDSs during natural disasters. Most recent literature has focused on developing model-based optimization approaches for effectively routing and scheduling MPSs to improve system resilience. For instance, authors in [8] applied a two-stage robust optimization approach to routing and scheduling of MPSs, aiming for a resilience-oriented outcome. In response to seismic events, the authors in [9] set up a two-stage mixed-integer nonlinear programming optimization model to optimize MPSs routing and scheduling for effective disaster recovery. A recovery strategy was introduced for PDSs that incorporates MPS deployment, addressing the variability in renewable energy sources through probabilistic constraints [10]. Authors in [11] tackled the issue of decision-dependent uncertainty related to MPS availability, influenced by travel and waiting times, providing a more accurate assessment of how MPSs can contribute to the enhancement of PDS resilience. Authors in [12] introduced a dynamic strategy for scheduling and routing of MPSs, taking into account the unpredictable condition of roads and electrical lines within integrated transport and power networks. Authors in [13] developed a multi-period mixed-integer linear programming co-optimization model that synchronizes the efforts of MPSs and RCs, aiming to enhance the resilience of PDSs. A co-optimization model that integrates the dispatch of MPSs and RCs through a mixed-integer second-order cone programming approach to fortify PDS resilience was proposed in [14]. Authors in [15] designed a service restoration model that not only increases the resilience of essential systems during disasters but also aligns the operations of MPSs with repair crew schedules, addressing limitations within both the power and transport sectors. Focusing on event prevention, Su et al. [16] advocate for a strategic public-safety power shutoff decision coupled with the deployment of MPSs. An innovative approach was introduced in [17], using battery-electric locomotives as mobile energy storage to manage wind energy variability and curtailment. However, the use of MPSs discussed in [8]–[17] rely solely on traditional energy sources, which produce harmful emissions and may be impacted by supply chain disruptions during disasters in rural areas. In contrast, MWTs are small-scale wind turbines designed for easy transport and are often used for off-grid power generation or powering remote locations [18]. The use of MWT fleets in rural PDSs has been extensively explored, including in energy management systems [19], pre-disaster management [20], and post-disaster restoration [21]. The existing literature [19]–[21] employs model-based optimization approaches, and due to concerns on computational complexity, all three studies use Monte Carlo simulations to address the uncertainty in wind power prediction represented by a limited representative number of scenarios.

AI-based data-driven approaches to addressing some of the computational challenges are being extensively studied. DRL is closely linked to optimal control and dynamic programming, offering significant advantages in real-time optimization of systems with imprecise or even in the absence of models (often known as model-free algorithms) [22]. DRL generates actions based on the current state, and following several training steps, it can determine the optimal actions for different states [23]. Due to these advantages, DRL has been applied in power systems management to enhance the resilience of power systems [24]. In response to the challenge of service restoration in PDSs during natural disasters, authors in [25] developed a decentralized MADRL framework for coordinated decision-making between MPSs and RCs aimed at enhancing resilience. In [26], a single-agent DRL method was proposed to make optimal dispatch decisions of MPSs for critical load restoration, accounting for uncertainties in electricity demand.

In [27], the authors developed a model-free real-time MADRL method for service restoration via routing and scheduling of MPSs in a coupled power-transport network.

B. Contributions and Paper Structure

To the best of our knowledge, previous studies [25]–[27] on utilizing DRL for MPS assignment have not examined the role and application of renewable-based MPS, such as MWT. In disaster-stricken regions, infrastructure damage can result in fuel shortages and road disruptions [6], [7], challenges that were not addressed in previous research. To bridge this gap, this article introduces a PDS restoration framework that incorporates MWT fleets. Unlike traditional MPSs, such as diesel generators, MWT fleets can continuously supply power without relying on fuel, but they also face unique challenges, particularly their dependence on wind availability, which requires strategic deployment to ensure reliable power generation. Road damage in rural areas adds further uncertainty, as it can significantly delay MWT fleet relocation. To address the impact caused by the wind speed uncertainty and provide a restoration strategy with time limitations, we propose a DRL-based resilience enhancement DRLBRE framework based on MADRL. In this framework, several key considerations can be highlighted including:

1) The proposed framework introduces a reward function that compares overall system rewards with individual agent rewards. This reward system helps the framework deliver higher-performing PDS restoration strategies using MWT fleets;
2) The framework incorporates action limitations to mitigate the impact of wind speed fluctuations, enhancing the stability of MWT fleet dispatch. This approach leads to improved training rewards by ensuring more reliable decision-making under varying wind conditions;
3) The framework applied a two-stage training process. In the first stage, a base model is trained on multiple fault scenarios in power-transport systems, establishing a solid foundation for the second stage, which refines strategies under specific faults caused by emergency natural disasters within time limitations. This approach minimizes training time limitations during emergency situations, enhancing the framework's efficiency, manageability, and ease of deployment.

The rest of this article is organized as follows: Section II provides an overview of the MWT technologies and the general framework, where the DQL and DDQL algorithms are also introduced; Section III provides the training reward and the numerical analysis of the studied cases; Section IV provides a summary of research findings and outlines prospects for future endeavors.

II. Problem Description

A. Mobile Wind Turbines

MWTs are small-scale wind turbines that are mounted on trailers or other mobile platforms, making them easily transportable to the desired locations to generate electricity [19]. Compared to backup power sources, such as uninterrupted power supply units used to restore the distribution system, MWTs offer the advantage of mobility. They can be transported by trucks to supply power to interrupted nodes at different locations, enabling continuous operation until all nodes are reconnected to the main grid. In addition, MWTs have a lower power generation cost compared to traditional diesel generators. Due to the fuel cost and the inefficiency of the diesel generator, the electrical power generation cost is 1 $\$/kWh$, while the electrical power generation cost for MWTs ranges from 0.07 to 0.25 $\$/kWh$ [28].

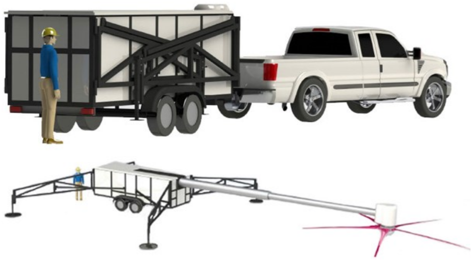


Fig. 1. A typical MWT setup [29].

Furthermore, there is a lack of research on service restoration schemes for rural PDSs, which face distinct challenges due to their isolated locations, constrained resources, and limited technical expertise. Natural disasters can severely disrupt rural supply chains, complicating the deployment of MPSs, especially diesel-based resources. Challenges in securing fuel and logistical difficulties post-disaster hinder timely power recovery efforts. Due to their independence from supply chains and their capability to generate green energy, MWTs emerge as a particularly effective strategy for swift service restoration in rural areas following extensive power outages. In this paper, we utilized MWTs from Uprise Energy [29], each capable of providing a maximum power output of 50 kW when connected to the PDS.

B. DRLBRE Framework for MWT-Enabled Restoration

The DRLBRE framework, devised to boost PDS resilience via MWT fleets, is depicted in Fig. 2. The framework is designed to rapidly generate a restoration strategy for the PDS using MWT fleets following a natural disaster, aiming to minimize economic losses. The strategy must account for faults in both the PDS and TS, as well as the inherent uncertainty of wind speed. To achieve a rapid response vital for effective restoration, we employ a DRL approach, which yields significantly faster decisions than traditional optimization methods [25]. Our framework is composed of two fundamental components: the environment and the DRL agent components. In the environment component, certain nodes of the PDS are linked to nodes in the TS, enabling MWT fleets to travel through the TS and supply power to PDS. In this study, we consider several PDSs integrated with a TS. When a disaster occurs, it results in damaged overhead power lines in PDSs and disconnected paths in the TS. The PDS and TS data, along with wind speed information, serve as input to the DRL agent model, constituting the model's state. This state inputs a DNN to determine actions using the ϵ-greedy function [30]. These actions guide the movement and power supply of

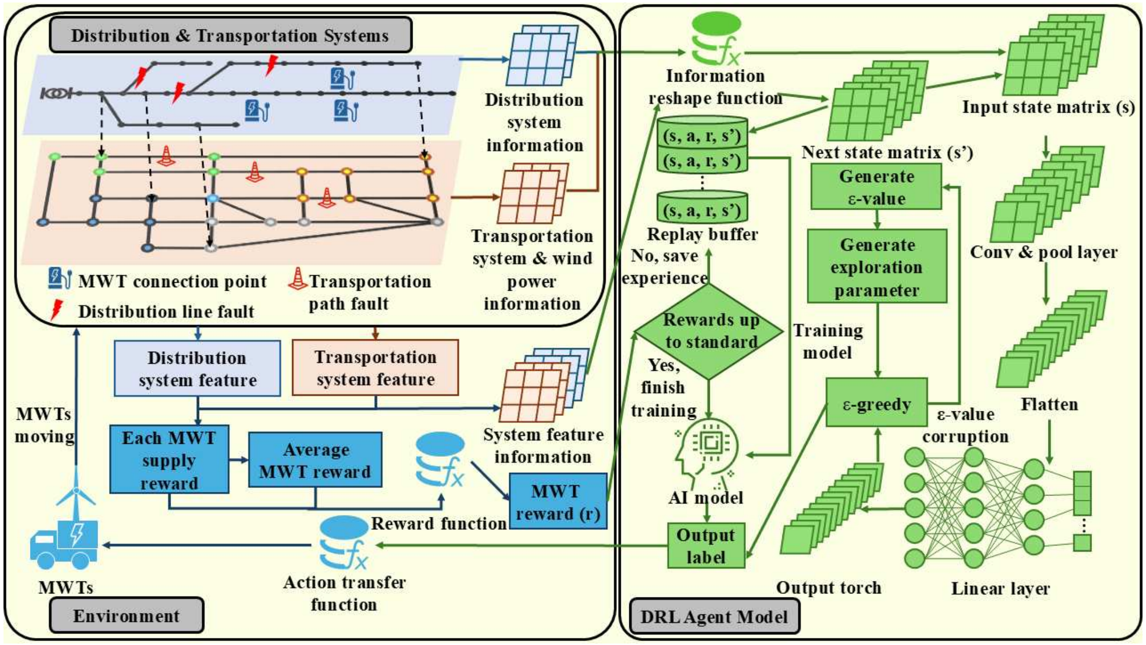


Fig. 2. The proposed framework for DRL-based resilience enhancement in PDS using MWTs.

the MWT fleets within the environment, following an action transfer function. The actions taken by agents lead to changes in environment information by altering the power generation of MWT fleets. A reward function evaluates the changes in power supply from MWT fleets, providing rewards to the DRL agent accordingly. The agent stores action, state, and reward data as experiences for training purposes. Training continues until the training reward meets a predefined standard, at which point model training ceases, and the model can generate actions based on the environment input states. To address the challenge of balancing the increased training time required by a complex environment with the time constraints imposed by emergency natural disasters, we adopted a two-stage training approach. In DRL, the two-stage method splits the training process into two phases, each designed with a specific focus [31]. This approach is commonly employed to handle complex tasks, beginning with the creation of a foundational model and then fine-tuning it for more specialized scenarios [32]. In our case, the first stage is dedicated to training the base model, during which PDS and TS faults along with wind power outputs, are randomly assigned for each training episode. The second stage involves application training, which uses well-defined configurations of PDS faults, TS faults, and random wind power conditions.

C. MADRL Algorithm

With the purpose of restoring the PDS following a natural disaster via MWT fleets, the DRL model is designed to provide action for each MWT fleet, which corresponds to selecting a target TN for power supply. The effectiveness of each action is evaluated only after its execution, as the travel time of MWT fleets can be different. In some cases, MWT fleets may need to continue supplying power at their current location based on the island demand and wind speed in TN. Differences in transportation time costs require that each MWT fleet follow an independent direction. Moreover, since multiple MWT fleets must collaborate to complete the restoration task, the DRL agent must consider not only the information from the PDS and TS, but also the status of all other MWT fleets, including their current locations and power output. This necessitates real-time information exchange and coordination among fleets to enhance both individual performance and the overall restoration strategy. Therefore, we implement MADRL in our framework. In this study, each MWT fleet applies one agent $A_i, i \in I$. The problem can be formulated as a MDP. All agents are DRL models, which enable agents to learn actions from their environments directly with exploration and exploitation [33]. A MDP is a 4-tuple (S, A, P, R), including $s \in S$ which stands for the state set of the environment; the agent action set A; P stands for the probability that s transfers to s' due to action a at time t; R represents the reward for taking action a at the current time step t. In each time step, the action $a_{i,t}$ is computed with an exploration policy conditioned on the current local state information $s_{i,t}$. The action $a_{i,t}$ is then applied to the environment, which responds by transitioning to a new state $s'_{i,t}$ and providing a reward $r_{i,t}$ for that action. Following this process, each agent i receives a local state information, action, reward, and the local state for the next time step as the experience $(s_{i,t}, a_{i,t}, r_{i,t}, s'_{i,t})$. The objective of each agent i is to maximize the total reward for the entire process $R = \sum_{t=0}^{T} r_{i,t}$, where T is the total time step.

In our framework, we applied DDQL and DQL as the DRL algorithm for each agent. The process for the MWT fleet agents is shown in Algorithm 1. The agent, denoted as i, belongs to the set I. The state for each agent at each time step consists of the information on the PDS, TS, power generation and MWT fleets location, shown as:

$$s_{i,t} = [N_{i,t}, P_w, f_P, f_T], \ i \forall I \tag{1}$$

$N_{i,t}$ is the transportation node information of the MWT fleet, P_w is the power generation of the MWT fleet, f_P is the fault information of PDS, and f_T is the fault information of the TS. The action of the agent $a_{i,t}, \forall i \in I$, represents the destination of the MWT fleet, indicating the TS node to which the MWT fleet should travel. Figure 3 illustrates how each agent generates actions and accesses information from other MWT fleets. When an MWT fleet completes its previous action or initiates the restoration process, its corresponding agent is activated and generates a new action based on the current input state, which is transferred from the environment information. This input state is composed of several matrices that include information about the PDS, TS, the current location of the MWT fleet, and the status of other MWT fleets. By modifying the input matrices, each agent gains experience in directing a specific MWT fleet. Once an action is completed, the environment updates and returns the state for the current time step. The information of the updated state triggers the agent to generate the next action, ensuring that each MWT fleet receives timely and individualized guidance for the restoration task. The reward $r_{i,t}, i \forall I$, stands for the cost of the load that is restored by the MWT fleets, which is used to evaluate the performance of the action. Further details will be discussed in Subsection III-A2.

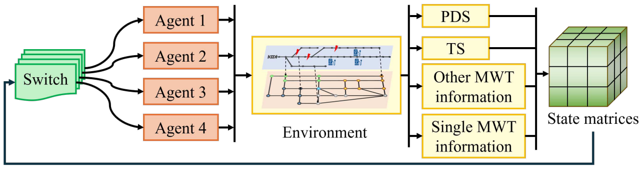


Fig. 3. Agent interaction cycle: information acquisition and action generation.

Algorithm 1 : DQL & DDQL Algorithm for MWT Agent

```
1:  Initialize replay memory $D_i$ $i \in I$ with capacity $N$
2:  Initialize the weight $\theta$ and $\theta^{'}$
3:  Set soft update of target parameters $\tau$ and leaning rate $\alpha_i$
4:  Set discount factor $\gamma$ and exploration rate $\epsilon$
5:  for episode $= 1, M$ do
6:      Set PDS and TS fault in environment
7:      Get the initial input state $s_{i,t}$ $i \in I$
8:      for time step $t = 1, T$ do
9:          for $i = 1, I$ do
10:             Generate action $a_{i,t}$ for each agent (MWT fleet) base on the $\epsilon$-greedy
11:             Transfer $a_{i,t}$ to the destination node $N_{i,d}$ of MWT
12:         end for
13:         Apply $N_{i,d}$ $i \in I$ in the environment
14:         Run the AC-OPF for PDS
15:         for $i = 1, I$ do
16:             Calculate reward $r_{i,t}$ with load restoration cost
17:             Get the next state $s^{'}_{i,t}$
18:             Store transition $(s_{i,t}, a_{i,t}, r_{i,t}, s^{'}_{i,t})$ in $D_i$
19:             Sample $(s_j, a_j, r_j, s^{'}_j)$ randomly from $D_i$
20:             Calculate the target value $y_j$ for DQL and DDQL
21:             Calculate the loss function for DQL and DDQL
22:             Update $\theta$ with loss function
23:             Update the state $s_{i,t} = s^{'}_{i,t} i \in I$
24:         end for
25:         Get episode-done value $e$ based on $t$
26:         if $e$ stand for current episode complete then
27:             Break the episode
28:         end if
29:     end for
30:     Decreases the $\epsilon$-value
31: end for
```

For the training process described in Algorithm 1, the initial steps involve setting up the hyperparameters and initializing the replay buffer, which stores the experiences for DQL and DDQL from Steps 1 to 4. At the start of each episode, in Step 6, faults in the PDS and TS are introduced to simulate a rural area affected by a natural disaster. In Step 7, the state $s_{i,t}$ for each agent is generated based on the conditions of the PDS and TS. From Steps 8 to 12, each agent uses the state information to determine its action following the ϵ-greedy policy. These actions are then translated into MWT fleet operations, including their power supply status and locations within the PDS and TS. In Steps 13 and 14, the actions of the MWT fleet are applied to the environment, and the AC-OPF is calculated for the PDS. Steps 16 to 18 involve each agent calculating the reward value based on the AC-OPF results, obtaining the next state $s'_{i,t}$, and storing the experience in their respective replay buffers. Finally, in Steps 19 and 20, the agents randomly sample training data from the replay buffer and calculate the target value y_j for DQL and DDQL. The target value for DQL is:

$$y_j = r_j + \gamma \max_{a'} \hat{Q}(s'_j, a'; \theta^-) \tag{2}$$

Note that DDQL uses two separate networks—an online network for action selection and a target network for target Q-value calculation—reducing overestimation bias, whereas DQL uses a single network for both, which can lead to overestimation. The target value of DDQL is calculated with the weight of the two Q networks:

$$y_j = r_j + \gamma \hat{Q}(s'_j, \mathrm{argmax}(Q(s'_j, a'))) \tag{3}$$

In Step 21, the loss function of the DQL and DDQL is calculated based on the target value as follows:

$$L_i(\theta_i) = (y_j - Q(s_j, a_j; \theta))^2, \ \ i \in I \tag{4}$$

In Step 22, the weight of the Q network in DQL and DDQL is updated with the learning rate α_i and discount parameter. For the Q network of DQL and local Q network of DDQL, the weight update equation is:

$$\theta_i \leftarrow \theta_i - \alpha \cdot \bigtriangledown_{\theta_i} L_i(\theta_i) \tag{5}$$

while the weight update equation for the target network of DDQL is:

$$\theta^{'}_i \leftarrow \tau \theta_i + (1 - \tau) \theta^{'}_i \tag{6}$$

In Step 23, the state information is updated using the state data in the next time step. From Steps 25 to 28, the process checks whether the cycle should end on the basis of the number of time steps. In Step 30, the exploration parameter ϵ decays accordingly.

D. AC-OPF in PDS

The restored load power in the PDS is calculated through an AC-OPF process. In each time step t, based on the MWT fleet power supply information $P^{wt}_{g,t}$, AC-OPF is used to quantify the restored load power. The objective is to maximize the expected cost sum of restored loads for PDS. The total restoration load cost is the sum of the restoration cost of each load d, which is calculated based on the load outage cost c_d^{lo} and the real power demand for each restored load $P_{d,t}^{ld}$ at each time step, formulated as

$$\left\{ \max_{\Xi^{mg}} \mathbb{E} \left\{ \sum_{d\in B_{ld}} \sum_{t\in T} c_d^{lo} P_{d,t}^{ld} \right\} \right. \tag{7}$$

subject to

$$\Xi^{mg} = \{P_{d,t}^{ld}, Q_{d,t}^{ld}, P_{g,t}^{mg}, Q_{g,t}^{mg}, P_{g,t}^{wt}, Q_{g,t}^{wt}, P_{ij,t}, Q_{ij,t}, V_{i,t}, V_{j,t}, e_{i,t}, y_{ij,t}\} \tag{8}$$

$$\sum_{g\in B_{wt}} P_{g,t}^{wt} + P_{g,t}^{mg} + = \sum_{d\in B_{ld}} P_{d,t}^{ld} - \sum_{(j,i)\in L} P_{ji,t} + \sum_{(i,j)\in L} P_{ij,t}, \forall i \in B, t \in T \tag{9}$$

$$\sum_{g\in B_{wt}} Q_{g,t}^{wt} + Q_{g,t}^{mg} + = \sum_{d\in B_{ld}} Q_{d,t}^{ld} - \sum_{(j,i)\in L} Q_{ji,t} + \sum_{(i,j)\in L} Q_{ij,t}, \forall i \in B, t \in T \tag{10}$$

$$P_{d,t}^{ld} \le e_{i,t} \bar{P}_{d,t}^{ld}, \forall i \in B, \forall d \in B_{ld}, t \in T \tag{11}$$

$$Q_{d,t}^{ld} \le e_{i,t} \bar{Q}_{d,t}^{ld}, \forall i \in B, \forall d \in B_{ld}, t \in T \tag{12}$$

$$e_{i,t} \underline{V}^2 \le V_{i,t}^2 \le e_{i,t} \bar{V}^2, \forall i \in B, t \in T \tag{13}$$

$$P_{ij,t}^2 + Q_{ij,t}^2 \le y_{ij,t} \cdot \bar{S}_{ij}, \forall (i,j) \in L, t \in T \tag{14}$$

$$V_{i,t}^2 - V_{j,t}^2 \le 2 \cdot (r_{ij} P_{ij,t} + x_{ij} Q_{ij,t}) + (1 - y_{ij,t}) \cdot M, \forall (i,j) \in L, t \in T \tag{15}$$

$$V_{i,t}^2 - V_{j,t}^2 \ge 2 \cdot (r_{ij} P_{ij,t} + x_{ij} Q_{ij,t}) + (y_{ijt} - 1) \cdot M, \forall (i,j) \in L, t \in T \tag{16}$$

The objective function (7) aims to maximize the cost sum of restored loads of the PDS. The formulated AC-OPF incorporates real and reactive power balance constraints (9) and (10) at bus i. The sets B_{wt}, B_{ld}, and L represent the MWT fleets, loads, and power distribution lines, respectively. The power supply from the main grid is denoted as $P_{g,t}^{mg}$. The status of demand at load point d is governed by constraints (11) and (12), while $e_{i,t}$ is a binary variable indicating the energized status of bus i (1 if energized, 0 otherwise). Voltage and power flow limitations for each bus and distribution line are described in (13) and (14), and the linearized power flow constraints are provided in (15) and (16). The binary variable $y_{ij,t}$ represents the energized status of distribution line (i,j) (1 if energized, 0 otherwise) and depends on the fault settings in the PDS. Since the distribution line undergoes repairs during the restoration process, $y_{ij,t}$ should be checked at each time step in the training episode. M represents a large positive value introduced to help simplify or relax the given constraints.

III. Numerical Case Studies

A. Environment Setup

The MADRL environment simulates the problem that needs to be addressed. The agent will make decisions based on the current state of the environment. The framework's environment encompasses four components: (i) a system simulation of the TS and PDSs; (ii) a wind speed generation function and MWTs output power estimation; (iii) the action transfer function of the MWTs, which converts the DRL agent's output into actions for the MWTs within the framework; (iv) an action reward function, which estimates rewards based on the power supplied by MWTs and the power outage costs at PDS load points.

1) The Integrated Power-Transport Network: Assuming that MWTs are used as assets aiding in the restoration of the rural PDS following disasters, we consider a limited-scale PDS. Therefore, we utilize four IEEE 33-bus test systems to represent the rural PDS. For the TS, we apply the Sioux Falls system, which consists of 24 TS nodes. For each PDS, there are candidate nodes coupling with nodes in TS. This allows MWTs to travel between coupling nodes in the TS, as determined by the DRL agent, to connect to the PDS and supply power. In this study, faults within the PDS are represented as broken vulnerable lines in PDSs, marked by red lightning to signify potential outages. The fault configurations within the PDS entail randomly selecting several power lines to be failed at different times throughout the process. This results in the segmentation of the PDS into multiple disconnected islands. Certain sections are isolated from the main grid and require power from MWTs. The complete PDS restoration process lasts 24 hours and is divided into 48 time steps. To model

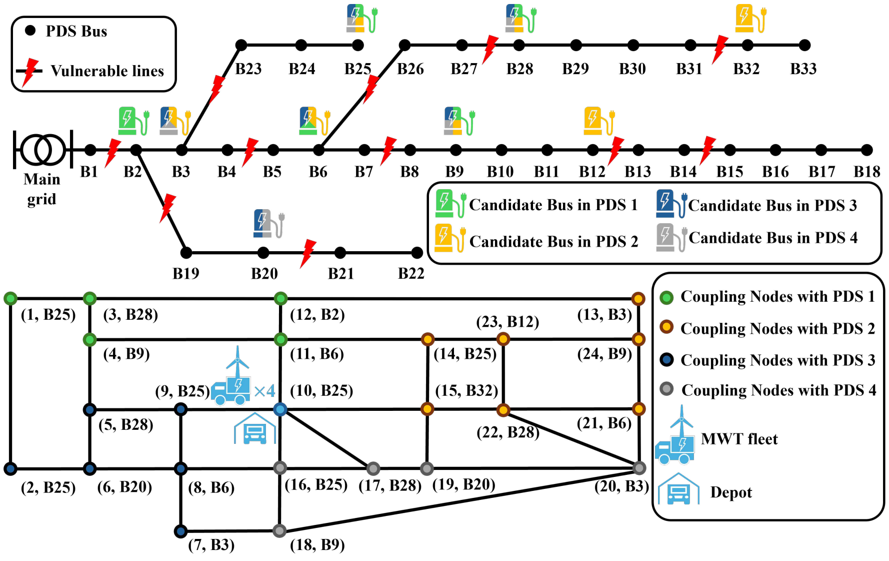


Fig. 4. An integrated power-transport network with IEEE 33-bus test system and Sioux Falls TS.

the fault scenario in the PDS, we randomly select four or five distribution lines to be damaged in each PDS. Simulating the actions of repair crews, we assume each line will be fixed after a certain number of time steps. The repair time is randomly assigned, ranging from 23 to 35 time steps.

The lower section of Fig. 4 depicts the TS, where nodes of different colors indicate the coupling relationship between the TS and each PDS. The TS fault is modeled as road damage caused by a disaster. Throughout the environment's process, these damaged roads will not be repaired, maintaining the existing road conditions to restrict the transport of MWTs. In each case within the environment, three roads in the TS will be randomly selected to be unavailable during the disaster.

To enable MWTs to navigate the TS using the shortest path, we used Dijkstra's algorithm to calculate the optimal route [34]. When considering faults in the TS, damaged roads are excluded when running Dijkstra's algorithm. During training, when the MWT receives a destination node—derived from the action generated by the DRL agents—the environment provides routing information based on the MWT's current TS node location and the destination TS node. The MWT then moves along the calculated shortest path until it reaches its destination. Since TS road damage in rural areas cannot be repaired quickly, the shortest path information remains unchanged for an entire training episode, which represents the whole restoration process. To reduce training computation and prevent recalculating the shortest path for each agent, we pre-compute a shortest path matrix at the beginning of each training episode.

2) Wind Speed and MWT Output Power: The power output of a wind turbine follows a truncated cubic relationship with wind speed [35]. In this context, the Weibull distribution is commonly employed due to its effectiveness in modeling uncorrelated wind speeds [36]. Since wind speeds vary across different times and locations, we use the Weibull distribution to randomly generate wind speed values for each TN at every time step in each episode. Given that the proposed framework is intended for post-disaster deployment, the MWT fleets are likely to operate under adverse conditions, including elevated wind speeds commonly observed during extreme weather events such as hurricanes. When wind speeds exceed the cut-out speed, the MWT will stop power generation as a protective measure to protect wind turbines from potential damage. The wind speed data is generated prior to the start of each episode. Since the same MWT is used for all units, the power output of each wind turbine is determined for every TS node at each time step using the generated wind speed data. The wind turbine power output function is given by:

$$P = \begin{cases} \frac{v^3}{v_{max}^3} P_{max} & 0 < v < v_{max} \\ P_{max} & v > v_{max} \\ 0 & v > v_{out} \end{cases} \tag{17}$$

P represents the wind power output of the MWT; v is the wind speed at the current time step; v_{max} is the rated wind speed of the wind turbine; P_{max} is the rated wind power output of the wind turbine; and v_{out} is the cut-out speed of the wind turbine. Since the primary task of the MWT fleet is to efficiently restore PDS nodes by supplying power, the MWT fleet's power generation capability is critical for evaluating performance and calculating rewards for the DRL agent. Low or zero power generation from the MWT fleet results in a lower reward. During training, the reward mechanism encourages the agent to avoid actions that position the MWT fleet in areas with low wind speeds or winds exceeding the cut-out speed. The MWT considered in this paper is capable of generating power over a wide range of wind speeds [29], with maximum power output achieved at a wind speed of $11m/s$ [37]. The cut-out speed is set to $25m/s$ [38]. The wind power information for the next time step is provided as input to demonstrate that the agent receives predictive wind power data. The MWT focused in this paper has the ability to generate power with a wide range of wind speeds [29]. Each MWT has a maximum power output of 50 kW. GFM converters have been implemented in microgrids and islanded power systems [39], enabling wind power plants to operate similarly to traditional synchronous power stations. By adjusting current and voltage, these converters help manage interconnection fluctuations [40]. Consequently, we deployed multiple MWTs and grouped them into a fleet, establishing a wind power plant capable of aiding restoration efforts. The maximum power output for each MWT fleet is limited to 1000 kW. Figure 5 illustrates a wind power output scenario for MWTs. The TS consists of 24 nodes, and the entire process spans 48 time steps, resulting in a 24×48 matrix. In this study, we deploy four MWT fleets with identical wind turbine power output configurations. Each colored block in Fig. 5 represents the wind power output of a single MWT fleet operating at a specific TS node during the corresponding time step.

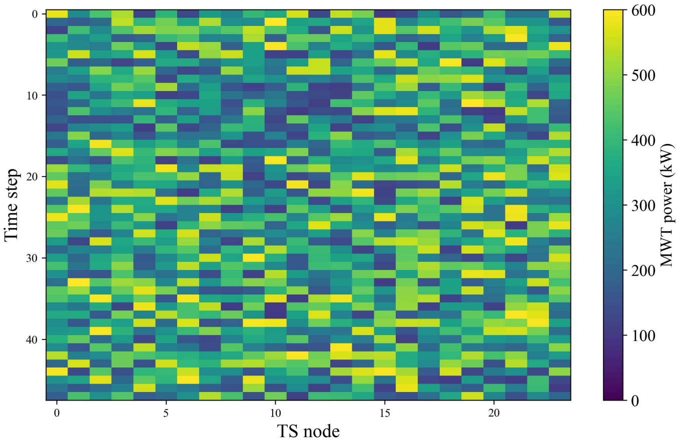


Fig. 5. MWT wind power output at each time step in PDS.

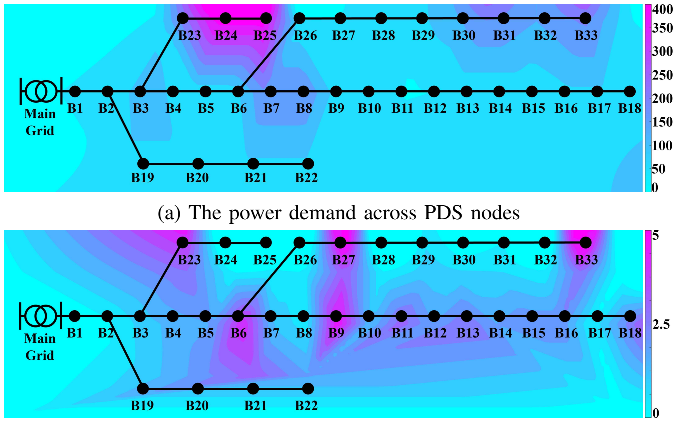
3) Action and Transfer Function: The restoration scheduling policy for the PDS comprises a series of actions, each corresponding to a specific destination TN for MWT fleets. These actions are represented by output labels from the DRL agent, and the total number of labels defines the action space size. In our simulation environment, there are 23 labels, aligning with the number of TNs, excluding the TN where the depot is located. The action transfer function is used to transfer the action label information to the action of the MWT fleets, which is the destination TNs of the MWT fleets. The inherent uncertainty in wind speeds, coupled with potential faults in PDS and TS, significantly expands the state space within the MDP framework. When combined with an extensive action space, this leads to a substantial increase in the number of potential state-action pairs, or MDP tuples. Such an expansion poses challenges for DRL agents, as the enlarged and more complex state-action space complicates the exploration process, making it more difficult for agents to identify optimal restoration actions for each state. To enhance exploration efficiency within our DRL framework, we implement an AL function that constrains the action space by excluding choices likely to result in suboptimal performance. Specifically, the AL function prevents scenarios where multiple MWT fleets simultaneously supply power to a single TN. This restriction mitigates the compounded negative effects of low wind speeds by preventing multiple MWT fleets from simultaneously delivering insufficient power to the PDS. Additionally, the AL mechanism helps avoid potential power waste that can occur when multiple MWT fleets generate power in the same TN under high wind speed conditions, thereby preserving the overall operational efficiency of the fleet. By narrowing the action space to more promising options, the AL function facilitates more effective exploration and accelerates the convergence of the DRL agents toward optimal restoration strategies. The algorithm for the action transfer function is described in Algorithm 2.

Algorithm 2 : Agent Action Transfer Function

```
Getting the action label of other agent and storage to $A_{list}$
Getting the current agent action number $a$
Getting the depot TN number $n_{depot}$
Getting the action number lower bound $a_{lower}$ and upper bound $a_{upper}$ based on action setting
Getting the agent location node number $n_{location}$
Combining the $A_{list}$, $n_{depot}$, $a_{lower}$, and $a_{upper}$ as the list of prohibited nodes $N_{prohibit}$
for node number $n_i$ in $N_{prohibit}$ do
    if $a >= n_i$ and $a < n_{i+1}$ then
        Change the action $a$ as $a_T$ according to $a_T = a + i$
        Break the For loop
    end if
end for
if $a_T = a_{upper}$ then
    Change the action $a_T$ as $n_{location}$ to make the MWT fleet stop to supply power
end if
```

The Action Transfer Function algorithm involves three key steps: *Step (i)* it identifies the bounds of the action number, the actions of other agents, and the TN number of the depot to establish a list of prohibited nodes; *Step (ii)* it transfers the agent's action based on this list by comparing the action number and node number in the prohibited node list; *Step (iii)* it compares the transferred action with the action size to determine whether the agent should halt and supply power or not. Including the actions of other agents in both *Step (i)* and *Step (ii)* by adding them to the forbidden node list constitutes the AL aimed at preventing multiple MWT fleets from converging on the same node within the TS. This enables the AL to mitigate the influence of wind speed variations at individual TNs on the total power output of MWT fleets.

4) Reward Function Setting: The reward function serves as a pivotal aspect of the environment, providing an assessment of the feedback received from the agents' actions. Our objective is to optimize the episode reward throughout the restoration process following the disaster. The episode reward is calculated as the sum of the rewards at each time step. The reward is calculated based on the cost of the load restored by MWT fleets. Figure 6 illustrates the power demand and outage costs at each PDS node, where darker colors indicate higher demand and higher power outage costs. This paper posits that all four PDSs use the same information (load demand and power outage costs) for each node, as illustrated in Fig. 6.



(a) The power demand across PDS nodes

(b) The power outage cost at PDS load points

Fig. 6. The power demand and the power outage cost of PDS nodes

To ensure that MWT fleets do not supply power to the island connected to the main grid, their reward value is converted into a penalty to influence the target value of the MADRL. The objective is to optimize the episode reward, which is the aggregate of the rewards from four episode MWT fleets. Therefore, the reward for each agent at each time step should be considered in two different ways: one is to use the SR, which is the average of the rewards from the four agents, and the other is the AR. The SR influences the agent's target value based on the performance of all agents' actions, while the AR encourages the agent to concentrate on its own action performance. The AR obtained by each agent when the MWT fleet under its control is supplying power is determined by the following equation:

$$R_{MWT} = c_{load} \cdot \frac{p_{MWT}}{p_{total} N_f} \tag{18}$$

where, R_{MWT} represents the reward value for each agent, and c_{load} denotes the cost of the load supplied by the MWT in the current section of the PDS. p_{MWT} signifies the wind power output of the MWT fleet, while p_{total} indicates the total power supplied by the MWT fleet in the current section of the PDS. N_f represents the transfer coefficient, which diminishes the magnitude of the reward. Our framework prioritizes minimizing economic losses in the PDS and improving system resilience. This means that developing an effective power restoration strategy takes precedence over other considerations. Incorporating fuel costs into the reward calculation could skew the restoration strategy, preventing it from supplying power to the most critical loads and potentially lowering the restored load's overall value. Hence, transportation cost is recognized but treated as a secondary concern, ensuring that it does not interfere with the primary objective of rapid and effective power restoration. Furthermore, MWT fleet's fuel consumption is negligible compared to the total value of lost load.

B. DNN Model and State Generation

In this paper, we employ a CNN as the DNN part of the MARDRL agents. This CNN is responsible for classification, using input matrices to assign action labels. These matrices encapsulate information from our framework environment related to MWT fleets, PDSs, and TS. The types of features in the input matrices are detailed in Table I. These feature information will be transformed into five matrices as input for the CNN, with the feature data types presented in Table I.

TABLE I
ENVIRONMENT OUTPUT FEATURE INFORMATION IN INPUT MATRIX

Feature Type	State Matrix Number
TS Feature	1
PDS Fault	2
Wind Speed Feature	3
Other MWT fleets Status	4
Other MWT fleets Location	4
Other MWT fleets Supplying Power	4
Other MWT fleets Action	4
MWT fleet Status	5
MWT fleet Location	5
MWT fleet Supplying Power	5
MWT fleet Action	5

In Table I, State Matrix 1 contains feature information on the length of transportation paths following the occurrence of faults due to disaster-induced damage. State Matrix 2 furnishes information on the status of distribution lines, contributing to the state's understanding of the PDS. State Matrix 3 offers wind speed features for each TN. State Matrix 4 presents features of MWT fleets controlled by other agents, encompassing MWT fleet status, location, supplied power, and action values. State Matrix 5 provides feature information on the MWT fleet controlled by the current agent. State Matrices 1 to 3 offer general feature information, while the feature details provided by State Matrices 4 and 5 aid the agent in understanding its own situation and the status of other agents.

The CNN will provide the agent's action label based on the input state information. In CNN, there are two Convolutional (Conv) layers, two Pooling (Pool) layers, and three Fully Connected (FC) layers in total. The architecture of the proposed CNN is: **Input(5, 12×12) – Conv1(64, 10×10) – Pool1(64, 5×5) – Conv2(128, 4×4) – Pool2(128, 2×2) – FC1((64×2)×2×2, 144) – FC2(144, 72) – FC2(72, action size)**. The learning rate η is 5e-3, and the minimum exploration rate ϵ is 1e-3.

C. Base Model Training

This paper employs two RL networks: DQL and DDQL. Utilizing these networks, we investigate suitable reward strategies and action transfer functions within this framework by testing various algorithm combinations to identify the optimal configuration for training. We calculate the total reward by summing the episode rewards of four agents to assess training performance. The moving reward represents the average value of the total reward over the last 100 episodes. The moving reward curve is shown in Fig. 7. In the legend label of Fig. 2, SR means applying system reward during the training, AR means applying the agent reward during the training, and AL means applying the action limitation in the action transfer function. We utilize the episode number at which the network achieves a moving reward of 1,500 as the benchmark for evaluating the training speed. Details regarding the training speed are provided in Table II.

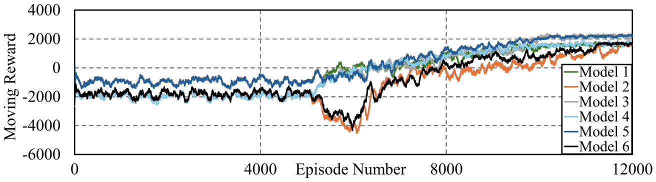


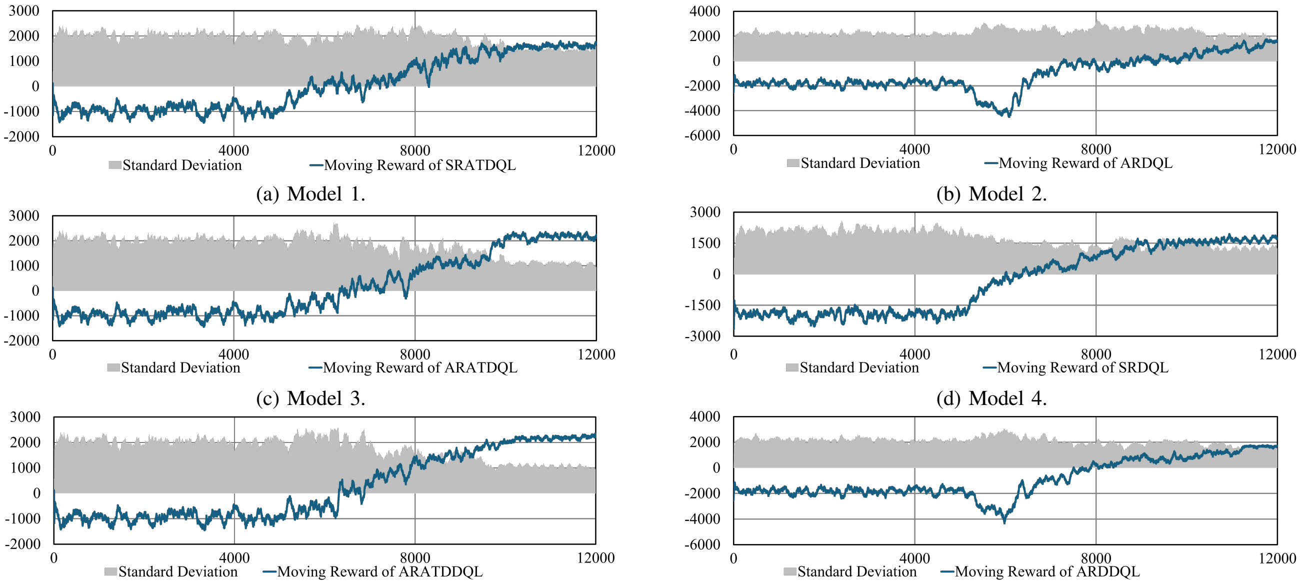
Fig. 7. Training moving reward for different networks.

TABLE II
TRAINING EPISODE OF DIFFERENT MODELS

Model Number	AR	SR	AL	DQL	DDQL	Episode Number
1		√	√	√		9,065
2	√			√		11,208
3	√		√	√		9,364
4		√		√		9,188
5	√		√		√	9,673
6	√				√	11,214

In Fig. 7, when comparing the moving reward curve of Model 2 and Model 4, it is evident that without the AL in the action transfer function, the SR outperforms the AR. The model using SR motivates agents to take actions that increase the total episode reward, while the model using AR prompts agents to choose actions that enhance their individual rewards. Comparing the moving reward curve with those of other models employing AR functions, it is evident that the episode moving reward rises faster. This acceleration is due to the system reward function, which allows individual agents to make sacrifices for the benefit of the overall reward. In the absence of AL, since AR leads agents to disregard the rewards of others, training encourages agents to select the optimal action for themselves, which may result in multiple agents supplying the same island and thus not utilizing the maximum wind power generation. Consequently, agents operating under SR could achieve higher episode rewards during training and converge sooner. However, agents that sacrifice their AR value to achieve a higher total episode reward may lead the MADRL model to converge to a suboptimal solution. This occurs because actions with a higher AR are replaced by those with a higher SR but lower AR during the MADRL training process. Based on the training results, AL reduces episode costs by 1.3% to 13.7%. Additionally, the final moving reward of training with AR and AL is over 31.8% higher than that of training with SR and AL. Considering the need for the framework to deliver strategies as quickly as possible, we incorporate AR and AL into the MADRL framework.

Upon incorporating AL with SR in Model 1, the episode cost of increasing the moving reward is lower compared to that in Model 4. Similarly, when comparing Models 2 and 3, as well as Models 5 and 6, within each pair of models sharing the same DRL structure and reward function, it becomes evident that AL significantly accelerates the training speed. After implementing AL, agents will assign different actions to each MWT fleet, directing them to various destination TNs. Since



(a) Model 1.

(b) Model 2.

(c) Model 3.

(d) Model 4.

(e) Model 5.

(f) Model 6.

Fig. 8. The moving reward and the episode reward standard deviation for models with different combinations.

wind speed at each TN is randomly generated, the reward for the same action will vary with changes in wind power. In the absence of AL, actions by the agents can result in multiple MWT fleets supplying power to the same TN. This concentration on a single node can lead to greater fluctuations in wind speed and reduce the overall episode reward. The agent with SR allows some cases to achieve a higher episode reward, which means that the restoration policy is more efficient for the cases of the current episode. However, the experience from these episodes may cause some agents' replay buffers to accumulate experiences with high system rewards but low individual agent rewards. When there are changes in fault settings and wind power, these agents lack experiences with high individual rewards, which can reduce the total reward for the episode. This defect leads to the final moving reward of Model 1 being lower than that of Model 3. Compared to the moving reward curve of Models 3 and 5, the training speed and the moving reward do not have obvious differences. Due to the random fault settings in both the PDS and TS during the disaster, the reward for each episode fluctuates based on these settings, resulting in a non-zero standard deviation in Fig. 8. To compare the performance of Model 3 and Model 5, we use these models as the base model for training with the same fault setting.

D. Case Model Training

In our framework, the second stage of training, known as case model training, is focused on developing a restoration strategy for the PDS to handle emergency natural disasters. This stage is designed to produce a time-sensitive restoration solution that effectively meets emergency challenges within a limited training period. To enhance performance and speed up the training process, we utilized the model trained during the first stage as the foundation for the second stage. For the case model training, each training episode used a single fixed fault setting. In our test scenarios, the PDS fault configuration is detailed in Table III, while the TS fault setting involves three fault paths: $6-8$, $8-16$, and $14-23$. The reward curves for various agent model configurations are presented in Fig. 9. In the second stage of training, the MADRL agents retain the same structure as used during the base training phase. To compare and evaluate the effectiveness of our two-stage training approach, we conducted multiple additional training runs using identical agent configurations. These additional runs utilized only the replay buffer from the first stage, without loading the network weights of the base model. In this second stage of training, we adjusted the value of ϵ to enhance exploration, helping agents identify actions that perform better for the current scenario. Two initial ϵ-values were considered: (i) 1.0, to encourage greater exploration of models without prior experience, and (ii) 0.15. Furthermore, since models 3 and 5 in Fig. 7 exhibit similar levels of training reward, we incorporated both into the case training for further comparison. These models were treated as base models for the case training phase to determine which could deliver a more effective restoration strategy.

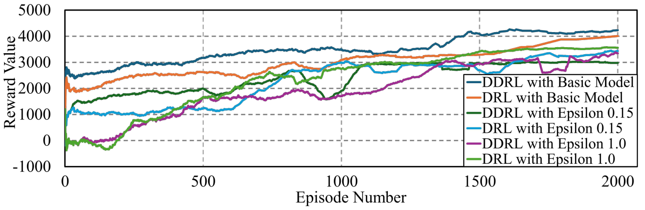
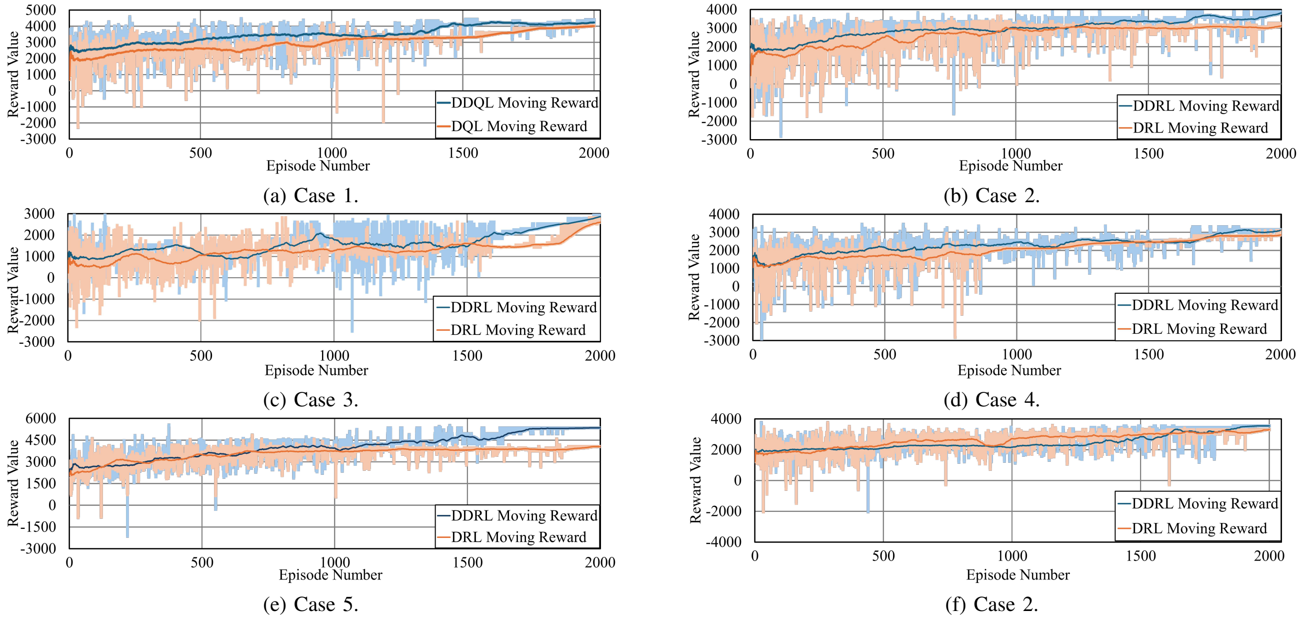


Fig. 9. Moving reward during model training with varying initial model settings.

In Fig. 9, training sessions utilizing base models, trained with various fault cases in the environment, attain higher moving rewards. Due to the sufficient experience possessed by the base model, high-reward actions can be achieved from the beginning of the training process. This prior experience enables rapid acquisition of high-reward actions. Consequently, the base model enhances the efficiency of developing transport strategies for MWT fleets. Comparing the two trainings with base models, the agent's performance using the DDQL net-



(a) Case 1.

(b) Case 2.

(c) Case 3.

(d) Case 4.

(e) Case 5.

(f) Case 2.

Fig. 10. The training moving reward in different cases.

TABLE III
FAULT SETTING IN FOUR PDSS IN CASE 1

PDS #	**Fault Line Endpoint Node 1**	**Fault Line Endpoint Node 2**	**Fault time step**
I	2	3	24
	3	23	34
	6	26	28
II	1	2	33
	7	8	23
	2	19	33
	6	26	34
III	4	5	28
	14	15	34
	2	19	30
	20	21	23
	31	32	31
IV	1	2	26
	7	8	34
	6	26	25
	27	28	30
	31	32	25

work outperforms that of the agent using the DQL network. The final moving reward of the DDQL network is 5.5% higher than that of the DQL network, and the DDQL network's moving reward converges earlier than the DQL network's. This is because DDQL helps avoid the excessive optimism in Q-value estimation seen with DQL, allowing the agent to reduce training costs per episode and achieve higher rewards through more frequent Q-value iterations. However, since the base model already possesses substantial experience from initial training, the agent with the DDQL network can secure higher rewards consistently from the beginning to the end of the training in a specific environment setting.

Additionally, the final performance of training without the base models and using an epsilon value of 1.0 exceeds that of training with an epsilon value of 0.15. The higher epsilon value prolongs the exploration phase, allowing the training to cover more episodes and accumulate more experience. This increases the likelihood of achieving better actions and rewards, particularly for agents with an epsilon value of 1.0. When comparing the training without the base model, the agent utilizing the DQL network secures better moving rewards by the end of the training process than those using the DDQL network. The caution of DDQL in avoiding excessive optimism means that 2,000 episodes may not be adequate to iterate the network towards achieving high moving rewards. Thus, the performance between DQL and DDQL presents a contrast to that observed in agents trained with the base model. Figure 10 illustrates various cases depicting the training moving reward curve and the disparity in different fault settings in TSs and PDSs. In various scenarios with different environment settings, the agent's moving reward consistently demonstrates the superiority of the DDQL model. Once training converges, the DDQL model consistently achieves higher rewards compared to the DQL model.

TABLE IV
TRAINING EPISODE TIME COST

Basic Model	**Initial ϵ Value**	**DDQL**	**DQL**	**Episode Cost (s.)**
√	0.15	√		2.79015
√	0.15		√	2.69349
	0.15	√		2.91075
	0.15		√	2.72304
	1.00	√		2.90691
	1.00		√	2.72928

Table IV displays the training time cost for each episode. When comparing episodes with identical base models and ϵ-values, training with the DDQL network requires more time than the DQL network due to the differences in the target loss values. Using the base training model reduces the time required for each episode. The CPU for training is $i7-12700$, and the GPU is $NVIDIA-3060Ti$. Based on our previous training comparisons of reward functions, DRL algorithms, and action transfer functions, we have opted for AR as the reward setting for our environment. Each agent utilizes the DDQL algorithm, and we apply AL in the action transfer function to maximize rewards during training.

As indicated by the training results in Figure 9 and Table

IV, our two-stage training approach enhances the restoration strategy's performance within defined constraints. This framework allows training to occur before natural disasters strike, enabling the development of time-sensitive restoration strategies to effectively address emergency challenges. By leveraging pre-training during the initial stage, our two-stage method improves emergency response performance when a disaster occurs. This approach mitigates the impact of limited training time during emergencies, making the framework more efficient, straightforward to manage, and easier to implement.

E. MWT Fleets Restoration Strategy

To demonstrate the restoration strategy derived from the proposed framework, we present the MWT fleets restoration strategy based on the training results from Case 1. The environment fault setting in Case 1 is identical to that used in the training, as shown in Fig. 9. Figure 11 provides detailed information on the fault settings and restoration strategy of the MWT fleets. The upper section shows the fault locations in the distribution lines across four PDSs, while the lower section illustrates the transportation road faults in the TS and the MWT fleets' actions for movement and power supply during the restoration process. The PDSs fault duration information is provided in Table III.

In Fig. 11, MWT fleet 1 first moves to TN 4 and supplies power to PDS 1 from Bus 9 between time steps 4 and 23. Once the distribution line between Bus 2 and Bus 3 in PDS 1 is repaired, the island supplied by MWT fleet 1 is reconnected to the main grid. MWT fleet 1 then moves to TN 6 of the TS to supply power to PDS 2 until all distribution line faults are repaired. MWT fleet 2 moved to supply power in PDS 2, initially providing power to TN 15 from Bus 32. After the distribution line fault in PDS 2 between Bus 7 and Bus 8 is repaired, MWT fleet 2 moves to TN 23 and supplies power from Bus 12 in PDS 2. According to the load information in Fig 6, the repair of the distribution line makes the island containing Bus 12 significantly more valuable than the one with Bus 32, as the new island has a higher load demand and greater output value. MWT fleet 3 first moves to TN 3 and supplies power from Bus 32 in PDS 1 between time steps 7 and 20. With the PDS 1 distribution line faults between Bus 2 and Bus 3, and between Bus 6 and Bus 26 scheduled to be repaired at time step 28, MWT fleet 3 relocates to TN 1 at time step 24 to supply power from Bus 25 until all distribution line faults are repaired. MWT fleet 4 moves directly to TN 18 and continues supplying power to Bus 9 in PDS 4. The island containing Bus 9 in PDS 4 remains disconnected from the main grid until all distribution line faults are repaired. According to Fig. 6, the high outage cost due to the load demand ensures that the agent receives a significant reward.

Figure 12 illustrates the power contribution of each MWT fleet and compares the total supplied power to the four PDSs with and without the restoration strategy. The restoration strategies are generated by the trained model shown in Fig. 10. Comparing the green and the black lines in each case, it is evident that the trained models effectively supply power to the loads and reduce power outages at each time step in the PDSs. In each case, MWT fleets provided 8% to 10% of the load power. The reward function guides the agent to direct MWT fleets to supply power to higher-priority loads, reducing economic losses on the PDS load side by 12% to 16%.

F. Development Training

To address the temporal and spatial variability in MWT fleet power output, we incorporated additional HSS units as backup power sources. The HSS unit can store electrical energy and supply power to the PDS during emergency events. It consists of a water electrolyzer, hydrogen storage, and a fuel cell. The system stores electrical energy by electrolyzing water to produce hydrogen, which is later utilized by the fuel cell to generate power. Figure 13 illustrates the operational statuses of the MWT fleets, HSS units, and PDS, with each section representing different power supply scenarios. Part 1 shows that when the load and HSS units are connected to the main grid, the main grid supplies power to the load and charges the HSS units. Part 2 describes the situation when the island is disconnected from the main grid and no MWT fleet is supplying power, in which case the HSS units will provide power if their storage tanks contain hydrogen. Parts 3 and 4 depict the collaboration between MWT fleets and HSS units, where both supply power to the load, and the power output of the HSS units depends on the output of the MWT fleets and the island's load demand. Parts 5 and 6 illustrate scenarios where the MWT fleets generate sufficient power for the load. If the MWT fleets produce excess power beyond the HSS units' demand and the storage tanks are not full, the surplus

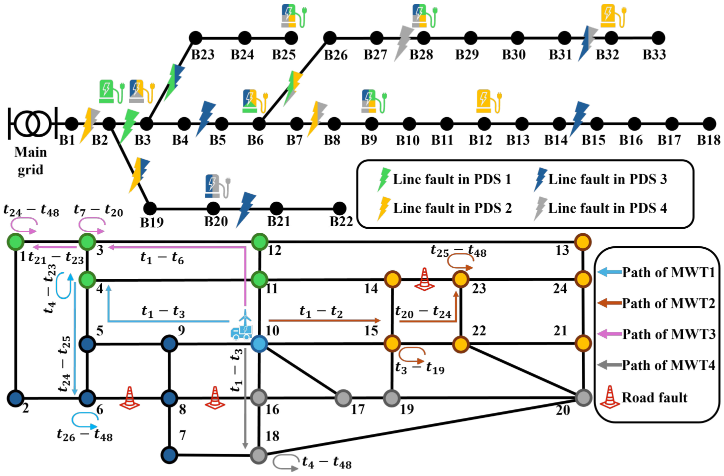
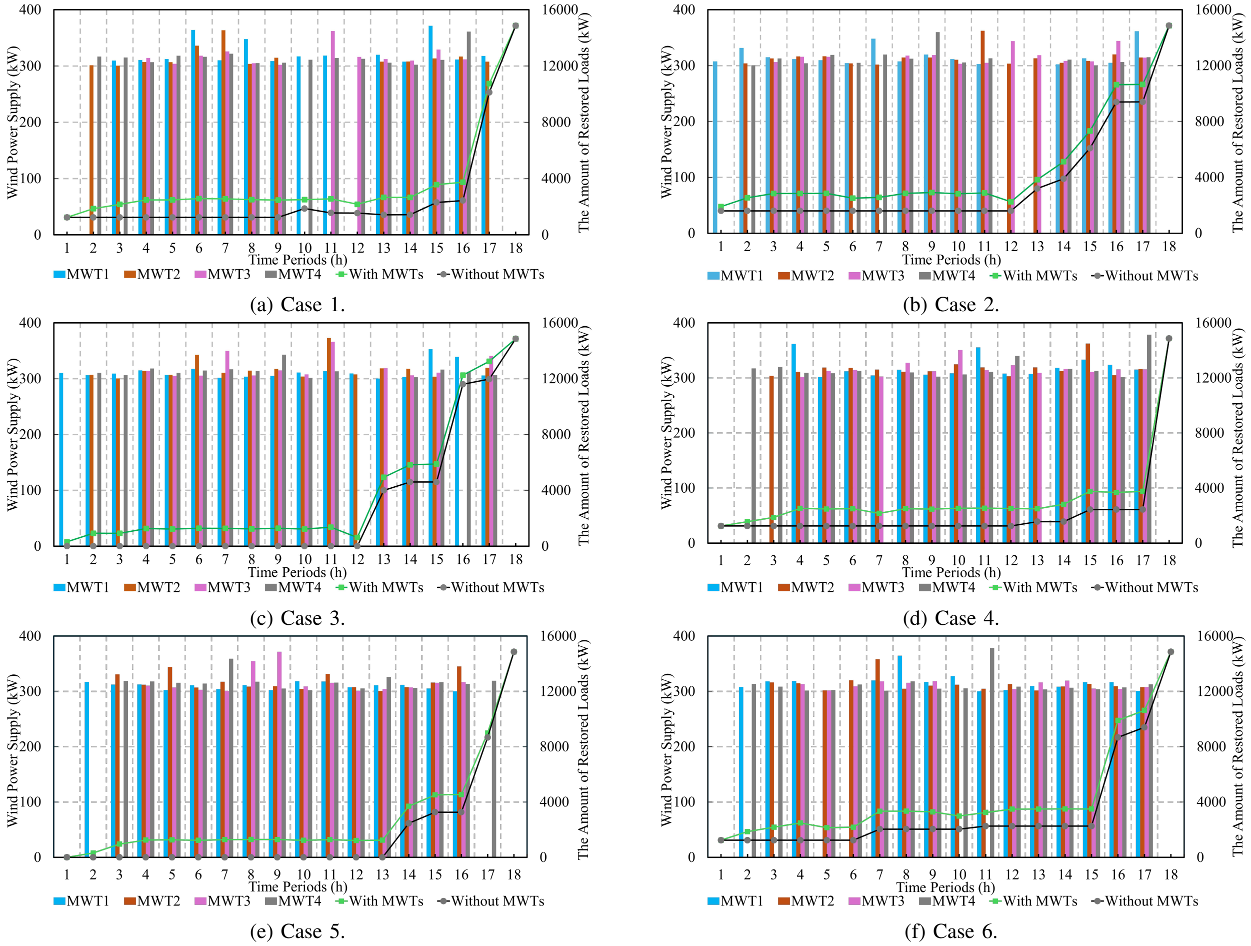


Fig. 11. The fault setting and the MWT fleets restoration action in Case 1.



(a) Case 1.

(b) Case 2.

(c) Case 3.

(d) Case 4.

(e) Case 5.

(f) Case 6.

Fig. 12. The assignment of MWT fleets in different cases.

energy is used to charge the HSS units, with the stored energy available for use when the MWT fleets cannot supply power to the island. Four HSS units were added to the PDS in Buses 11, 19, 23, and 26, as shown in Fig. 14. Based on commercially available products, we consider HSS units with a nominal power output of $100kW$ [41] in this study. The HSS units can be charged using an electrolyzer when surplus power is available on the island [42]. For the hydrogen storage tanks, we assume they have sufficient capacity to store hydrogen capable of generating a total of $200kWh$ of energy. The power output of HSS unit and the variations in the energy storage level of HSS unit are formulated as follows:

$$P_i^{h2p} = p_i^{h2p} \mu_i^{h2p} \eta_i^{h2p} \tag{19}$$

$$E_i^H = E_{i-1}^H - P_i^{h2p} \tag{20}$$

where P_i^{h2p} represents the power output of a single HSS unit, p_i^{h2p} indicates the hydrogen consumption by the HSS unit, μ_i^{h2p} denotes the H2P conversion factor, and η_i^{h2p} is the H2P efficiency of the HSS unit. E_i^H and E_{i-1}^H refer to the energy storage levels of HSS unit at the current and previous time steps, respectively. At the start of each training episode, the HSS unit begins with its storage fully charged and can supply power to islands disconnected from the main grid until the stored energy is exhausted. If MWT fleet generates more power than the island load demands and the connected HSS unit still has storage capacity available, the excess MWT fleet power is used to recharge the HSS unit. The relationships governing the charging power and energy storage of HSS unit are described by the following equations.

$$P_i^{p2h} = \frac{p_i^{p2h}}{\mu_i^{p2h} \eta_i^{p2h}} \tag{21}$$

$$E_i^H = E_{i-1}^H + P_i^{p2h} \tag{22}$$

where P_i^{p2h} is the hydrogen power input to a single HSS unit; p_i^{p2h} is the supply power from PDS to HSS unit; μ_i^{p2h} is P2H conversion factor of the HSS unit; η_i^{p2h} is the P2H efficiency of the HSS unit.

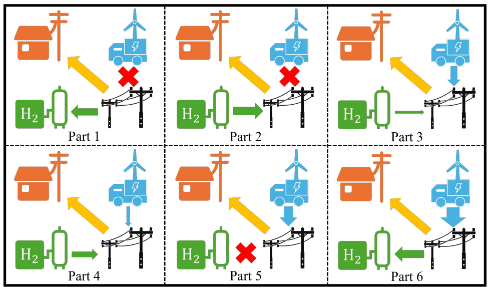


Fig. 13. PDS restoration status with HSS units and MWT fleets.

In Subsection III-C, we showed that integrating AL and AR enhanced training performance, enabling the framework to produce more effective restoration strategies. Compared to the agent model using the DQL algorithm, the agent model utilizing the DDQL algorithm further improved the overall performance of the framework. To further benchmark our approach against other established methods, we substituted the DDQL algorithm in our MADRL agents with three alternative

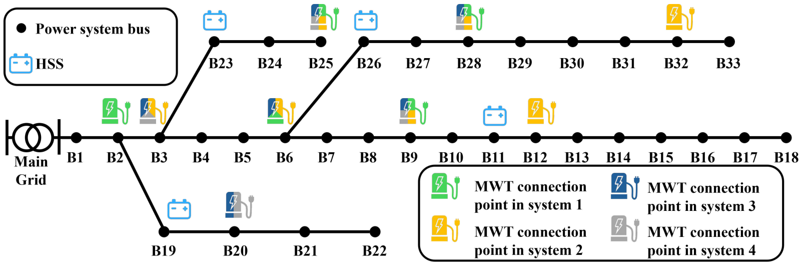

Power system bus
HSS
Main Grid
B1
B2
B3
B4
B5
B6
B7
B8
B9
B10
B11
B12
B13
B14
B15
B16
B17
B18
B19
B20
B21
B22
B23
B24
B25
B26
B27
B28
B29
B30
B31
B32
B33
MWT connection point in system 1
MWT connection point in system 2
MWT connection point in system 3
MWT connection point in system 4

Fig. 14. The modified PDS integrated with HSS units.

DRL algorithms: AC algorithm, DDPG algorithm, IPPO algorithm. We then evaluated the performance of the framework in terms of reward and time cost. Each agent used the AL and AR, and the training reward curves are presented in Fig. 15. The details of the converged rewards and the episode time costs are shown in Table V.

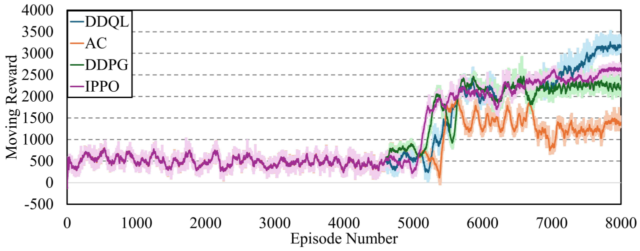


Fig. 15. The training reward for different DRL algorithms.

TABLE V
THE CONVERGE REWARD AND EPISODE COST FOR DIFFERENT DRL MODELS IN THE DRLBRE FRAMEWORK

Model Name	Converge Reward	Episode Time(s)	Converge Episode
AC	1408	2.76589	7868
DDPG	2268	3.25455	7105
IPPO	2685	2.78373	7655
DDQL	3248	2.79016	7685

Based on the training results in Table V, the training episode time cost for the frameworks using AC, IPPO, and DDQL is similar, while the framework with DDPG agents requires more time per episode. This is because DDPG consists of two AC networks, resulting in a total of four DNNs per agent, whereas the other three algorithms have only two DNNs per agent. The additional DNNs in DDPG increase the time required for agents to compute actions and calculate target values during training. As illustrated in Fig. 15, the framework that employs agents based on DDQL achieves notably higher rewards than those that use other DRL algorithms. This difference arises from the action space design for agents in the environment. Agents utilizing algorithms like AC, DDPG, and IPPO, which are based on PG methods, are well-suited for continuous action spaces due to the activation functions in their networks. In these algorithms, the agent action space is limited by the activation functions of the network to a continuous range between -1 and 1. However, in our research, the agent action corresponds to a target TS node, which must be selected from a discrete set of 24 values. To address this, we partition the continuous action space into 24 equal intervals, each corresponding to a specific TS node number. These small intervals make it more challenging for the agent to distinguish subtle differences in action values, increasing the probability of the agent selecting incorrect actions for restoration. In previous research [25], the authors utilized Proximal Policy Optimization, a PG algorithm. To address the challenge of continuous agent action spaces, the authors choose to establish a model using two agents that separate the direct moving information of a single MPS and the information of the power supply. One group of agents provides data on possible movement paths from the current TS node, while another set supplies details on the MPS power status. This method can decompose a single complex action into two simple actions and use two agents to provide two-action information. However, more agents mean that the model needs more networks and will increase the training time cost. With the limitation of the training time caused by the emergency natural disaster, we use a single agent to direct the MWTs action in our framework. Compared to the TS used in previous research [25], our TS is significantly more complex. This complexity leads to much smaller intervals when using a PG-based algorithm. Therefore, we opted to use the DDQL algorithm in our agents. The DDQL algorithm, being a value-based approach, naturally aligns with discrete action spaces, eliminating the need for action space conversions from continuous to discrete. Consequently, our framework, utilizing DDQL agents, achieves higher rewards and can address the restoration problem in more complex TS scenarios with reduced time costs.

TABLE VI
AVERAGE REWARD FOR 30 TEST CASES USING DIFFERENT MODELS

Method	AC	DDPG	IPPO	DDQL	MILP
Average reward	960.16	2384.57	2653.48	3193.696	3566.303
Computation (s)	83.1	97.5	83.4	83.7	2169.3

After obtaining the initially trained model, we use the DRL agent for testing and compare its performance with a traditional MILP-based optimization approach. The test is conducted across 30 distinct fault scenarios with randomized wind speeds to simulate rapid response conditions following a natural disaster. The results, presented in Table VI, are evaluated using the agent's reward function, which assesses the effectiveness of each mitigation action step by step. The comparison shows that our framework, which integrates AL and AR with DDQL, significantly enhances response speed while maintaining competitive performance. Figure 16 illustrates the reward values across 30 test cases, representing the performance of models employing different algorithms. The purple line indicates the performance achieved by the MILP-based optimization approach. Compared to the MILP model, our framework, which implements DDQL, provides equivalent restoration performance in 10 test cases. Among the frameworks utilizing various DRL algorithms, our approach achieves the best performance in 16 test cases. In contrast, the AC algorithm provides the best restoration strategy in only one test case, the DDPG algorithm leads in 6 test cases, and the IPPO algorithm performs best in 7 test cases. Our framework, leveraging DDQL, outperforms frameworks that implement other DRL algorithms, achieving the highest performance in the majority of test scenarios. Additionally, it demonstrates stable and competitive results compared to the traditional MILP optimization method, underscoring its effectiveness in diverse fault conditions within power-transport systems.

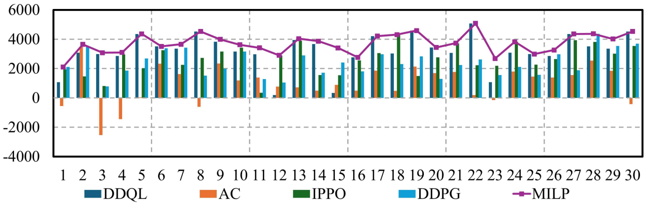
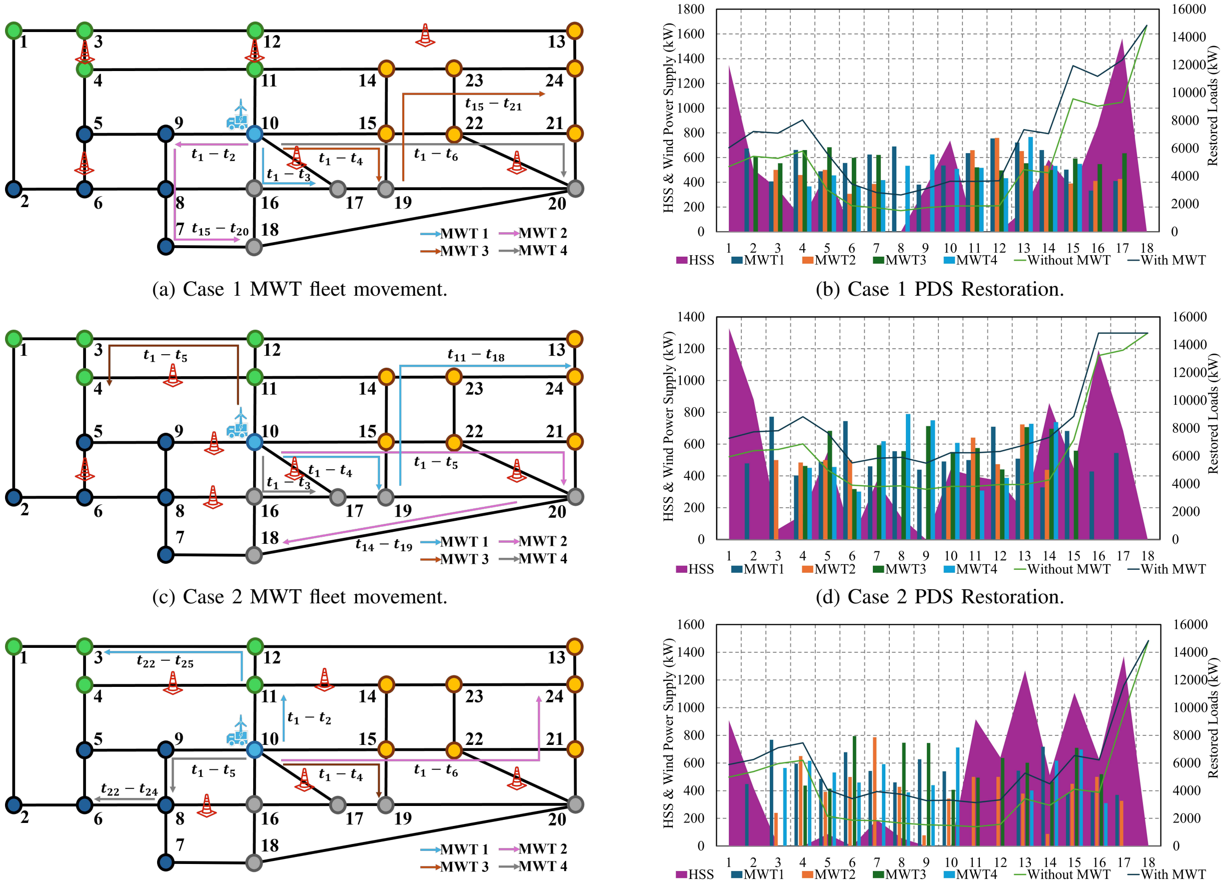


Fig. 16. The training reward curve for different DRL algorithms.

Figure 17 shows the results of three different test cases for PDS restoration with MWT fleets and HSS units. The left section of Fig.17 illustrates the movement of MWT fleets within the TS affected by road damage. The right section presents the power output of each MWT fleet, the total power output of the HSS units, and the system load restoration information. Using the information in the left section of Fig.17, our framework can generate a restoration strategy to identify the shortest path free from road damage. Based on the HSS units configuration in our PDS, each HSS unit can deliver $100kW$ power for 2 hours, with a total maximum power output of $1600kW$. When comparing the total power output of HSS units in three scenarios, with MWT fleet charging and the primary energy storage in HSS units, the HSS units can sustain the power supply to the PDS for a longer duration. During hours 8, 9, and 10 in Fig. 17(b), MWT fleet 2 and MWT fleet 3 are in motion within the TS and are unable to supply power to the PDS. As a result, the power supply from HSS unit increases, preventing a decline in PDS restoration power compared to hour 7. This demonstrates that HSS unit enhances the performance of the PDS restoration strategy in our framework. The inclusion of HSS unit also highlights the scalability of our framework in integrating additional energy resources.

IV. Conclusion

This paper introduces an innovative MADRL framework to tackle the coordinated dispatch of MWT fleets for PDS service restoration following extreme disasters. The proposed MADRL framework consists of two training stages: initially, the base model is trained using a variety of random environmental scenarios. Subsequently, this model undergoes further training with specific scenario settings to develop a scheduling strategy tailored to the current scenario. The MADRL neural network in the framework offers discrete actions with sensible transfer functions to devise scheduling restoration strategies



(a) Case 1 MWT fleet movement.

(b) Case 1 PDS Restoration.

(c) Case 2 MWT fleet movement.

(d) Case 2 PDS Restoration.

(e) Case 3 MWT fleet movement.

(f) Case 3 PDS Restoration.

Fig. 17. The test case result of restoration with MWT fleets and HSS units

for PDSs with MWT fleets, ensuring the supply of power to loads isolated from the main grid. To develop a more efficient restoration policy and shorten training time for emergency restoration tasks, we implement a DDQL network to establish the agent within the framework. We integrate AL into the action transfer function to accelerate the training process, reducing training episodes by 1.3% to 13.7%. Additionally, AR is used as the reward function for each agent, improving the training reward by 31.8% and enhancing the restoration policy. Numerical analyses with six different fault scenarios in a power-transport system, consisting of four IEEE 33-bus test systems and one Sioux Falls TS, demonstrated that the framework can provide efficient restoration strategies, supply power to loads separated from the main grid, and enhance the resilience of the PDS. The results from various cases with different power-transport system fault settings show that the restoration policy provided by our framework directs MWT fleets to supply 8% to 10% of the PDS load power. Additionally, our framework takes load value into account, guiding the MWTs to prioritize high-value loads within limited power, reducing economic losses by 12% to 16%.

References

[1] P. Dehghanian, B. Zhang, T. Dokic, and M. Kezunovic, “Predictive risk analytics for weather-resilient operation of electric power systems,” *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 3–15, 2019.

[2] “Hurricane irma and maria in puerto rico – building performance observations, recommendations, and technical guidance,” 2020. [Online] Available at: https://www.fema.gov/sites/default/files/2020-07/mat-report_hurricane-irma-maria-puerto-rico_2.pdf [Accessed: Jun. 12, 2025].

[3] J. W. Busby, K. Baker, M. D. Bazilian, A. Q. Gilbert, E. Grubert, V. Rai, J. D. Rhodes, S. Shidore, C. A. Smith, and M. E. Webber, “Cascading risks: Understanding the 2021 winter blackout in Texas,” *Energy Research & Social Science*, vol. 77, pp. 1–10, 2021.

[4] T. Ding, M. Qu, Z. Wang, B. Chen, C. Chen, and M. Shahidehpour, “Power system resilience enhancement in typhoons using a three-stage day-ahead unit commitment,” *IEEE Transactions on Smart Grid*, vol. 12, no. 3, pp. 2153–2164, 2021.

[5] J. Wassmer, B. Merz, and N. Marwan, “Resilience of transportation infrastructure networks to road failures,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 34, no. 1, pp. 1–18, 2024.

[6] E. Lindner, “More Than 400 Roads Closed in North Carolina After Damage From Helene,” 2024. [Online] Available at: https://www.nytimes.com/2024/09/28/us/roads-closed-western-north-carolina.html.

[7] S. Khan, “Florida Fuel Suppliers Brace for Shortages as Hurricane Helene Approaches,” 2024. [Online] Available at: https://www.usnews.com/news/us/articles/2024-09-26/.

[8] S. Lei, C. Chen, H. Zhou, and Y. Hou, “Routing and scheduling of mobile power sources for distribution system resilience enhancement,” *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5650–5662, 2019.

[9] Z. Yang, P. Dehghanian, and M. Nazemi, “Seismic-resilient electric power distribution systems: Harnessing the mobility of power sources,” *IEEE Trans. Ind. Appl*, vol. 56, no. 3, pp. 2304–2313, 2020.

[10] M. Nazemi, P. Dehghanian, X. Lu, and C. Chen, “Uncertainty-aware deployment of mobile energy storage systems for distribution grid resilience,” *IEEE Trans Smart Grid*, vol. 12, no. 4, pp. 3200–3214, 2021.

[11] J. Su, D. Anokhin, P. Dehghanian, and M. A. Lejeune, “On the use of mobile power sources in distribution networks under endogenous uncertainty,” *IEEE Trans. Control Netw. Syst.*, vol. 10, no. 4, pp. 1937–1949, 2023.

[12] S. Yao, P. Wang, X. Liu, H. Zhang, and T. Zhao, “Rolling optimization of mobile energy storage fleets for resilient service restoration,” *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1030–1043, 2020.

[13] T. Ding, Z. Wang, W. Jia, B. Chen, C. Chen, and M. Shahidehpour, “Multiperiod distribution system restoration with routing repair crews, mobile electric vehicles, and soft-open-point networked microgrids,” *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 4795–4808, 2020.

[14] S. Lei, C. Chen, Y. Li, and Y. Hou, “Resilient disaster recovery logistics of distribution systems: Co-optimize service restoration with repair crew and mobile power source dispatch,” *IEEE Transactions on Smart Grid*, vol. 10, no. 6, pp. 6187–6202, 2019.

[15] D. Anokhin, P. Dehghanian, M. A. Lejeune, and J. Su, “Mobility-as-a-service for resilience delivery in power distribution systems,” *Production and Operations Management*, vol. 30, no. 8, pp. 2492–2521, 2021.

[16] J. Su, S. Mehrani, P. Dehghanian, and M. A. Lejeune, “Quasi second-order stochastic dominance model for balancing wildfire risks and power outages due to proactive public safety de-energizations,” *IEEE Transactions on Power Systems*, vol. 39, no. 2, pp. 2528–2542, 2024.

[17] F. Kochakkashani, P. Dehghanian, and M. A. Lejeune, “On the use of battery-electric locomotive as a grid-support service in electric power systems,” *IEEE Transactions on Power Systems*, pp. 1–13, 2024.

[18] G. Erdemir, A. E. Kuzucuoğlu, and F. A. Selçuk, “A mobile wind turbine design for emergencies in rural areas,” *Renewable Energy*, vol. 166, pp. 9–19, 2020.

[19] J. Su, P. Dehghanian, B. Vergara, and M. H. Kapourchali, “An energy management system for joint operation of small-scale wind turbines and electric thermal storage in isolated microgrids,” in *2021 North American Power Symposium (NAPS)*, pp. 1–6, 2021.

[20] J. Su, R. Zhang, P. Dehghanian, and M. H. Kapourchali, “Pre-disaster allocation of mobile renewable-powered resilience-delivery sources in power distribution networks,” in *2023 North American Power Symposium (NAPS)*, pp. 1–6, 2023.

[21] J. Su, R. Zhang, P. Dehghanian, M. H. Kapourchali, S. Choi, and Z. Ding, “Renewable-dominated mobility-as-a-service framework for resilience delivery in hydrogen-accommodated microgrids,” *Int j. electr. power energy syst.*, vol. 159, p. 110047, 2024.

[22] J. Zhang, L. Sang, Y. Xu, and H. Sun, “Networked multiagent-based safe reinforcement learning for low-carbon demand management in distribution networks,” *IEEE Trans Sustain Energy*, vol. 15, no. 3, pp. 1528–1545, 2024.

[23] M. Zhang, G. Guo, S. Magnússon, R. C. N. Pilawa-Podgurski, and Q. Xu, “Data driven decentralized control of inverter based renewable energy sources using safe guaranteed multi-agent deep reinforcement learning,” *IEEE Trans Sustain Energy*, vol. 15, no. 2, pp. 1288–1299, 2024.

[24] M. M. Hosseini, L. Rodriguez-Garcia, and M. Parvania, “Hierarchical combination of deep reinforcement learning and quadratic programming for distribution system restoration,” *IEEE Trans Sustain Energy*, vol. 14, no. 2, pp. 1088–1098, 2023.

[25] Y. Wang, D. Qiu, F. Teng, and G. Strbac, “Towards microgrid resilience enhancement via mobile power sources and repair crews: A multi-agent reinforcement learning approach,” *IEEE Transactions on Power Systems*, vol. 39, no. 1, pp. 1329–1345, 2024.

[26] S. Yao, J. Gu, H. Zhang, P. Wang, X. Liu, and T. Zhao, “Resilient load restoration in microgrids considering mobile energy storage fleets: A deep reinforcement learning approach,” in *2020 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, 2020.

[27] Y. Wang, D. Qiu, and G. Strbac, “Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems,” *Applied Energy*, vol. 310, p. 118575, 2022.

[28] J. Knight, “Uprise Energy’s Mobile Wind Turbine: An Innovative Solution Showcased at TEVCON 2024,” 2024. [Online] Available at: https://upriseenergy.com/blog.

[29] Uprise Energy, “Portability,” 2024. [Online] Available at: https://upriseenergy.com/portability.

[30] J. Chen, L. Zhang, J. Riem, G. Adam, N. D. Bastian, and T. Lan, “Explainable Learning-Based Intrusion Detection Supported by Memristors,” in *2023 IEEE Conference on Artificial Intelligence (CAI)*, pp. 195–196, 2023.

[31] J. Wang, P. Zhang, and Y. Wang, “Autonomous target tracking of multi-uav: A two-stage deep reinforcement learning approach with expert experience,” *Applied Soft Computing*, vol. 145, p. 110604, 2023.

[32] J. Chen, H. Huang, Z. Zhang, and J. Wang, “Deep reinforcement learning with two-stage training strategy for practical electric vehicle routing problem with time windows,” in *International conference on parallel problem solving from nature*, pp. 356–370, Springer, 2022.

[33] Y. Mei, H. Zhou, T. Lan, G. Venkataramani, and P. Wei, “MAC-PO: Multi-Agent Experience Replay via Collective Priority Optimization,” in *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, pp. 466–475, 2023.

[34] W. Shu-Xi, “The improved dijkstra’s shortest path algorithm and its application,” *Procedia Engineering*, vol. 29, pp. 1186–1190, 2012.

[35] A. T. Abolude and W. Zhou, “Assessment and performance evaluation of a wind turbine power output,” *Energies*, vol. 11, no. 8, p. 1992, 2018.

[36] P. Dehghanian and M. Kezunovic, “Probabilistic decision making for the bulk power system optimal topology control,” *IEEE Transactions on Smart Grid*, vol. 7, no. 4, pp. 2071–2081, 2016.

[37] “Exceptional efficiency charged with smart programming results in unmatched performance.,” 2025. [Online] Available at: https://upriseenergy.com/performance.

[38] “How do wind turbines survive severe weather and storms?,” 2024. [Online] Available at: https://www.energy.gov/eere/articles/how-do-wind-turbines-survive-severe-weather-and-storms.

[39] R. Rosso, X. Wang, M. Liserre, X. Lu, and S. Engelken, “Grid-forming converters: Control approaches, grid-synchronization, and future trends—a review,” *IEEE open j. ind. appl.*, vol. 2, pp. 93–109, 2021.

[40] “What is Grid Forming?,” 2025. [Online] Available at: https://www.neso.energy/energy-101/electricity-explained/how-do-we-balance-grid/what-grid-forming.

[41] “100kw hydrogen fuel cell,” 2025. [Online] Available at: https://www.hydroxcel.com/100kw-hydrogen-fuel-cell/.

[42] “Water electrolysis system - 100 kw,” 2025. [Online] Available at: https://www.fuelcellstore.com/water-electrolysis-system-100kw.

Ruotan Zhang (Graduate Student Member, IEEE) received his B.Eng degree in electrical engineering and the automatization specialty from the Hunan University in 2018, and the M.Sc. degree in electrical engineering from The George Washington University, Washington, D.C., USA in 2021. He is currently pursuing a Ph.D. degree in Electrical Engineering at the Department of Electrical and Computer Engineering, The George Washington University, Washington, D.C., USA. His research interests encompass the mitigation of electromagnetic pulse (EMP) strikes in power systems, machine learning–assisted resilience enhancement for power systems, and the restoration of power distribution systems using renewable energy sources.

Jinshun Su (Graduate Student Member, IEEE) received B.Eng in electrical engineering from Xi’an University of Technology, China in 2017, and the M.Sc. degree in electrical engineering from The George Washington University, Washington, D.C., USA in 2019. He is currently pursuing the Ph.D. degree in electrical engineering at the Department of Electrical and Computer Engineering, The George Washington University, Washington, D.C., USA. His research interests include applications of mobile power sources for resilient smart grids, modeling and analysis of decision-dependent uncertainties in energy systems, solutions for optimal public-safety power-shutoffs to build resilience against electrically-induced wildfires

Payman Dehghanian (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Tehran, Tehran, Iran, in 2009, the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, in 2011, and the Ph.D. degree in electrical engineering from Texas A&M University, College Station, TX, USA, in 2017. In 2018, he joined the Department of Electrical and Computer Engineering, The George Washington University, Washington, DC, USA, where he is currently an Associate Professor. His research interests include power systems reliability and resilience assessment, data-informed decision making in power and energy systems, and smart electricity grid applications.

He was the recipient of the 2014 and 2015 IEEE Region 5 Outstanding Professional Achievement Awards, 2015 IEEE-HKN Outstanding Young Professional Award, 2021 Early Career Award from the Washington Academy of Sciences, 2022 George Washington University’s Early Career Researcher Award, 2022 IEEE IAS Electric Safety Committee’s Young Professional Achievement Award, and 2022 IEEE IAS Outstanding Young Member Service Award. In 2015 and 2016, he was selected among the World’s Top 20 Young Scholars for Next Generation of Researchers in electric power systems.

Mohannad Alhazmi (Member, IEEE) is an Assistant Professor at the Department of Electrical Engineering at King Saud University, Riyadh, Saudi Arabia. He received the B.Sc. degree in Electrical Engineering from Umm AlQura University, Saudi Arabia, in 2013, the M.Sc. degree in Electrical Engineering from The George Washington University, Washington D.C., USA, in 2017, and the Ph.D. degree in Electrical Engineering from The George Washington University, Washington D.C., USA, in 2022. His research interests include power system reliability and resiliency, as well as the operation of interdependent critical infrastructures. Dr. Alhazmi was a recipient of the 2022 IEEE Industry Application Society (IAS) Electrical Safety Prevention through Design Student Education Initiative Award.

Xiaoyuan Fan (Senior Member, IEEE) is a Regional Technology Manager in the Grid Intelligence group at Eaton Research Labs, Golden, CO, USA. Prior to joining Eaton, he was a Senior Staff Engineer and Power Electronics Team Leader with the Pacific Northwest National Laboratory, Richland, WA, USA. His research interests include data analytics for grid reliability, wireless communication, multi-discipline resilience analysis, and high-performance computing. He is a Senior Member of IEEE, and also a volunteer reviewer of more than 20 top-level journals and conferences in the area of power systems and signal processing. In addition, he is the recipient of two Eaton Corporation E-Star Awards, 2024 R&D 100 Award, 2024 Secretary of Energy’s Honor Award - Achievement Award, 2021 Federal Laboratory Consortium Award, and four PNNL Outstanding Performance Awards. Xiaoyuan received the B.S. and M.S. degrees in electrical engineering from the Huazhong University of Sciences & Technology, Wuhan, China, in 2009 and 2012, respectively, and the Ph.D. degree in electrical engineering from the University of Wyoming, Laramie, WY, USA, in 2016.