A Novel Reinforcement-Learning-Based Compensation Strategy for DMPC-Based Day-Ahead Energy Management of Shipboard Power Systems

Fu, Jianfeng; Sun, Dingshan; Peyghami, Saeed; Blaabjerg, Frede

DOI
10.1109/TSG.2024.3382213

Publication date
2024

Document Version
Submitted manuscript

Published in
IEEE Transactions on Smart Grid

Citation (APA)
Fu, J., Sun, D., Peyghami, S., & Blaabjerg, F. (2024). A Novel Reinforcement-Learning-Based Compensation Strategy for DMPC-Based Day-Ahead Energy Management of Shipboard Power Systems. *IEEE Transactions on Smart Grid*, *15*(5), 4349-4363. https://doi.org/10.1109/TSG.2024.3382213

A novel reinforcement-learning-based compensation strategy for DMPC-based day-ahead Energy Management of Shipboard Power Systems

Jianfeng Fu, Dingshan Sun, Saeed Peyghami, *Senior Member, IEEE,* and Frede Blaabjerg, *Fellow, IEEE,*

***Abstract*—Distributed model predictive control (DMPC) has become a focus in the energy management of shipboard power systems due to its capabilities for privacy preservation, robustness, and distributing computing burdens to local processors. DMPC determines control actions in a distributed manner based on the predictions of system statuses. However, the performance of DMPC is affected by inaccurate predictions resulting from uncertain parameters in nominal prediction models. Particularly, these inaccuracies in predicting propulsion loads and solar panel generation powers can lead to power imbalances when implementing the control actions determined by DMPC. To address this challenge, this paper proposed a novel reinforcement learning compensated DMPC (RL-C-DMPC) to distributively compensate for the control actions determined by DMPC baseline control, thereby rectifying the power imbalances caused by uncertain parameters in nominal prediction models. A value-decomposition-network-based training and distributed testing mechanism is designed for our proposed RL-C-DMPC. Furthermore, a method for range selection of compensation rate is specifically proposed for the energy management of shipboard power systems. To validate the effectiveness of our proposed RL-C-DMPC, we conduct a comprehensive case study utilizing real-life voyage data and historical solar power generation data in the area of the voyage to build the environment for training and testing. By comparing power imbalances between DMPC and RL-C-DMPC, our results indicate significant reductions in power imbalances so that frequency stability can be better ensured. Furthermore, via the case study, we also evaluate the communication robustness of RL-C-DMPC.**

***Index Terms*—Uncertainties, reinforcement learning, distributed control frameworks, distributed model predictive control**

NOMENCLATURE

Acronyms

ADMM	Alternating direction method of multipliers
ARIMA	Autoregressive integrated moving average
DG	Diesel generator
DMPC	Distributed model predictive control
DQN	Deep Q Network
MPC	Model predictive control
RL	Reinforcement learning
RL-C-MPC	Reinforcement learning compensated model predictive control
RL-C-DMPC	Reinforcement learning compensated distributed model predictive control
SOC	State of charge
VDN	Value-decomposition network

Sets and Indices

$\mathcal{I}$	The set of zones (also the set of agents)
$\mathcal{K}_t$	The set of time steps of the prediction horizon starting from the current time step t

Parameters

K_i^{Q}	Estimated advance coefficient for obtaining the propulsion load of propeller i
$\tilde{K}_i^{\mathrm{Q}}$	Actual advance coefficient for obtaining the propulsion load of the propeller in Zone i
$n_{i,k}^{\mathrm{pro}}$	Revolution speed of the propeller in Zone i
N	Number of zones/agents
NP	Maximum number of iterations
$\overline{P}_i^{\mathrm{bat}}$	The maximum charging power of the battery in Zone i
$\underline{P}_i^{\mathrm{bat}}$	The minimum charging power of the battery in Zone i
$\overline{P}_i^{\mathrm{gen}}$	The maximum generation power of the diesel generator in Zone i
$\underline{P}_i^{\mathrm{gen}}$	The minimum generation power of the diesel generator in Zone i
$P_{i,k}^{\mathrm{pro}}$	The predicted propulsion load of the propeller in Zone i at time step k using the nominal prediction model
$P_{i,k}^{\mathrm{ser}}$	The service load of Zone i at time step k
$P_{i,k}^{\mathrm{sol}}$	The generation power of the solar panel in Zone i at time step k using the nominal prediction model
$\underline{R}_i^{\mathrm{gen}}$	The minimum ramping rate of the diesel generator in Zone i
$\overline{R}_i^{\mathrm{gen}}$	The maximum ramping rate of the diesel generator in Zone i
$\underline{S}_i^{\mathrm{bat}}$	The lower SOC boundary of the batch of batteries in Zone i
$\overline{S}_i^{\mathrm{bat}}$	The upper SOC boundary of the batch of batteries in Zone i
α_i	Square cost coefficient of diesel generator in Zone i

This work is supported by the "National Natural Science Foundation of China" under Grant U1910216 and the "Fundamental Research Funds for the Central Universities" under Grant 3132023108.

Jianfeng Fu is currently with the School of Electrical Engineering, Dalian University of Technology, Dalian, China. He was previously with the College of Marine Electrical Engineering, Dalian Maritime University, Dalian, China, (e-mail:j.fu-1@outlook.com).

Dingshan Sun is with the Delft Center for Systems and Control, Delft University of Technology, Delft, The Netherlands, (e-mail: D.Sun-1@tudelft.nl).

Saeed Peyghami and Frede Blaabjerg are with AAU Energy, Aalborg University, (e-mail: sap@energy.aau.dk, fbl@energy.aau.dk).

β_i Linear cost coefficient of diesel generator in Zone i

ϵ_k Random value for predictions of generation powers of the solar panel at time step k

η_i Comprehensive coefficient of environment and propeller parameters

ϕ_i Estimated parameters of the autoregressive process of the solar panel in Zone i

ρ Penalty parameter of the augmented Lagrangian term

$\tilde{\phi}_i$ Actual parameters of the autoregressive process of the solar panel in Zone i

θ_i Estimated parameters of the moving process of the solar panel in Zone i

$\tilde{\theta}_i$ Actual parameters of the moving process of the solar panel in Zone i

ξ_1, ξ_2 Tolerance gaps of synchronous ADMM

Variables

$P^{\mathrm{bat}}_{i,k}$ The charging power of the batch of batteries in Zone i at time step k

$P^{\mathrm{gen}}_{i,k}$ The generation power of the diesel generator in Zone i at time step k

$P^{\mathrm{int}}_{i,j,k}$ The interconnecting power flow from Zone i to Zone j at time step k

$S^{\mathrm{bat}}_{i,k}$ SOC of the batch of batteries in Zone i at time step k

Learning related parameters and variables

$P^{\mathrm{cp}}_{i,t}$ Compensation power rate of the generation powers of the diesel generator in Zone i at time step t

$\underline{P}^{\mathrm{cp}}_{i,t}$ Minimum compensation power rate of $P^{\mathrm{cp}}_{i,t}$

$\Delta P^{\mathrm{cp}}_{i,t}$ Linear increment of compensation power rate of $P^{\mathrm{cp}}_{i,t}$

$\hat{P}^{\mathrm{bat}}_{i,t}$ Baseline charging power of the batch of batteries in Zone i at time step t

$\hat{P}^{\mathrm{gen}}_{i,t}$ Baseline generation power of DG in Zone i at time step t

$\tilde{P}^{\mathrm{gen}}_{i,t}$ Compensated generation power of DG in Zone i at time step t

$\tilde{P}^{\mathrm{pro}}_{i,t}$ Actual propulsion load in Zone i at time step t

$\tilde{P}^{\mathrm{sol}}_{i,t}$ Actual generation power of the solar panel in Zone i at time step t

$u^{\mathrm{F}}_{i,t}$ Actions (outputs) of the DQN of the agent in Zone i at time step t

$v^{\mathrm{O}}_{i,t}$ System state of Zone i at time step t

$v^{\mathrm{P}}_{i,t}$ State for predictions of Zone i at time step t

$w^{\mathrm{P}}_{i,t}$ State (input) of the DQN in Zone i at time step t

γ Discount factor of training

σ Signal indicating whether the ramping boundaries are satisfied

I. Introduction

In recent years, there has been increasing attention and application in the maritime and transportation industries towards hybrid vessels that combine both conventional and renewable energy sources [1]–[3]. This shift is in response to the International Maritime Organization's commitment to reduce total greenhouse gas emissions from international shipping [4], [5]. As part of this trend, hybrid vessels, especially all-electric diesel-solar vessels, have gained considerable attention in both research and practical deployment [6]–[8]. However, due to the intermittent nature of uncertainties of propulsion loads and renewable energy generation power predictions, it becomes crucial to develop intelligent energy management strategies to ensure an optimal, safe, and stable operation of shipboard power systems.

In the literature, distributed model predictive control (DMPC) has been extensively studied as an energy management strategy for both shipboard and on-shore power systems because of its capabilities for privacy preservation, robustness, and distributing computing burdens to local processors [9]–[11]. The DMPC approach first involves separating the power systems into several local zones and formulating the local energy management optimization problems for zones. Then, these local energy management optimization problems are solved in a distributed manner by local agents to obtain the control actions for each zone. Subsequently, the local agents execute the control actions in a distributed manner. Being a model-based approach, DMPC relies on nominal prediction models to predict parameters within a prediction horizon. Consequently, the accuracy of these nominal prediction models significantly influences the overall control performances of DMPC, such as optimality.

One of the main factors affecting the accuracy of the nominal prediction models is uncertainties. Uncertainties lead to inaccurate parameter estimations of nominal prediction models that are biased from those of the actual prediction model. The uncertainties may come from the degradation of equipment and system [12], [13], the inaccurate measurement device [14], [15], or/and epistemic reasons [16], [17], etc. In the most context of day-ahead energy management for onshore and shipboard power systems, power balance should be strictly satisfied [18]–[20]. However, inaccurate predictions of propulsion loads and solar panel generation powers can result in power imbalances. Although primary and secondary regulations can handle a part of power imbalance in smaller time scales, their regulation capability may not be enough. Thus, these imbalances may pose a significant challenge to the frequency stability of shipboard power systems and underscore the critical importance of handling uncertainties in day-ahead energy management.

Accordingly, several reinforcement learning (RL)-based methodologies have been presented in the literature to address the challenge of uncertainties. The first methodology directly uses RL to determine control actions based on power system statuses [21]–[27]. However, a major challenge with this method lies in effectively handling hard constraints. The second methodology explores the use of RL to formulate more accurate nominal models that consider uncertainties [28]–[32]. These improved models are then applied in DMPC or MPC-based energy management strategies. However, as these models are formulated individually, it is challenging to ensure the overall performance of the entire system. Moreover,

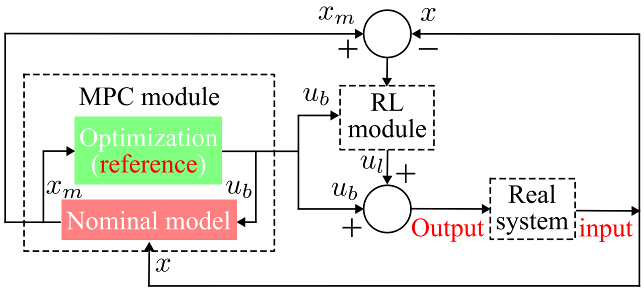


Fig. 1. The proposed control structure of using RL to compensate MPC in [35].

formulating accurate models that account for all uncertainty sources is impractical, particularly when some sources are even unknown and interact with each other.

Several recent studies are using RL to compensate for baseline control actions, such as back-stepping control and MPC, to handle uncertainties and to ensure overall performance. For example, in [33], [34], RL is employed to compensate for back-stepping control and synchronization control to avoid the collision of vessels and cable-driven robots, respectively. Additionally, one of the co-authors of this paper proposed to use RL to compensate for MPC in the domain of traffic control [35]. The control structure is shown in Fig. 1, where u_b and u_l are control actions obtained by MPC and the compensation value, respectively. Furthermore, x and x_m represent the actual states and the states derived by substituting u_b into the nominal model. The output of the control structure in [35] is the sum of u_l and u_b, and the inputs are the actual states feedback by the real system. The reference of the control structure is implemented as the objective function of the "optimization" section in the MPC module to maximize or minimize certain indices. Although the strategies proposed in [33]–[35] demonstrate that RL-compensated baseline control can effectively handle uncertainties and disturbances, it is essential to note that these works mainly focus on centralized control frameworks.

Furthermore, to the best knowledge of the authors, in the literature, no other papers apply compensation techniques (compensating the control actions obtained by baseline control strategies) to energy management of shipboard power systems. Some papers in the literature use compensation techniques in DMPC for microgrids [36]–[38]. However, these compensation techniques are not based on reinforcement learning. On the contrary, our paper proposes a reinforcement-learning-based compensation technique for DMPC.

The state-of-the-art RL-based compensation strategies designed for centralized control frameworks cannot be directly applied to distributed control frameworks due to the physical coupling of the local zones. To address this challenge, this paper introduces a novel RL-compensated DMPC (RL-C-DMPC) strategy tailored for distributed control frameworks. Given that each agent can only access local states from its respective zone, this paper designs a centralized training mechanism to train deep Q networks (DQNs) [39]–[41] based on value-decomposition-network (VDN) [42]–[44]. After training, the DQNs of the zones effectively compensate for control actions in a distributed manner. Furthermore, this paper proposes a method for selecting proper ranges of compensation specifically for the energy management of shipboard power systems.

The contributions of this paper are listed as follows:

- It proposes a novel RL-compensated DMPC (RL-C-DMPC) approach to use RL to distributively compensate for the control actions of DMPC. Firstly, the approach effectively addresses hard constraints, which present challenges for traditional RL-based decision-making methodologies. Secondly, it considers the overall performance of the system. Lastly, RL-C-DMPC avoids the need to formulate accurate models for all uncertainty sources, making it a practical and efficient solution.
- The research on using RL to compensate for the control actions obtained by DMPC in a distributed control framework has not been studied yet, so the proposed RL-C-DMPC is to explore this new area.
- This paper designs a value-decomposition-network-based training and distributed testing mechanism for the proposed RL-C-DMPC.
- It also introduces a method to select a proper compensation range specifically for the energy management of shipboard power systems. This method can reduce the scale of the action space and avoid insufficient compensation.
- The communication robustness of the proposed RL-C-DMPC is evaluated.

II. Preliminaries

A. DMPC-based day-ahead energy management of hybrid vessels

An illustrative layout of shipboard power systems of all-electric hybrid vessels is shown in Fig. 2. The layout in Fig. 2 is modified based on the crew-training and experimental ship "YuKun" of Dalian Maritime University [45]. The portrait, diesel generator, after-deck, and the propeller revolution speed measurement and control panel of "YuKun" are shown in Fig. 3. This paper assumes that three groups of solar panels are installed on the fore-deck, railing, and after-deck of "Yunkun", and three batches of batteries are installed.

The layout of Fig. 2 is divided into three zones that include several components in the shipboard power system. In shipboard power systems of hybrid vessels, this paper considers diesel generators that provide main powers, batteries that enable flexible operations, solar panels that generate renewable energy, service loads for crew and goods, and propulsion loads for propelling the vessels. To control the shipboard power system distributively, each zone is managed by an agent. The agents can communicate with each other via communication wires, determine, and execute the control actions of their zones distributively. The control actions of zones include the generation powers of the diesel generators and the discharging/charging powers of the batches of batteries.

At the beginning of each time step, agents collect the current states, predict future states using the nominal prediction models, determine the control actions for one prediction horizon by

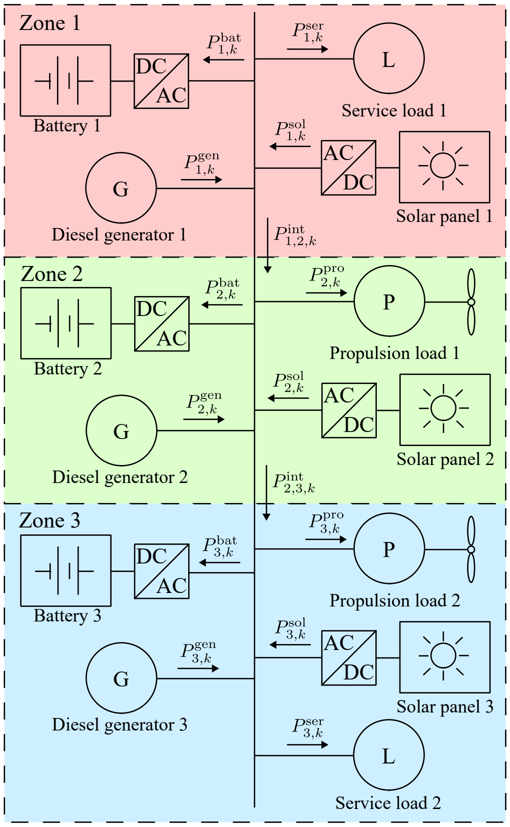
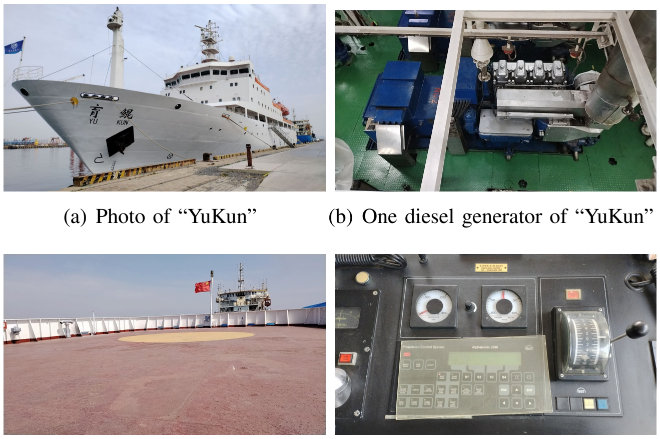


Fig. 2. An illustrative layout of shipboard power systems of all-electric hybrid vessels.



(a) Photo of "YuKun"

(b) One diesel generator of "YuKun"

(c) After-deck of "YuKun"

(d) Propeller revolution speed measurement panel of "YuKun"

Fig. 3. Portrait and main equipment of "YunKun".

solving the DMPC energy management optimization problem, and then only execute the control actions of the first time step according to the receding mechanism of DMPC.

B. Two sources of uncertainties in nominal prediction models

DMPC determines control actions using the nominal prediction model. However, because of uncertainties, the parameters of the actual prediction models cannot be accurately obtained or measured in practice. Thus, the parameters of the nominal prediction models used in the DMPC strategy may be biased from those of the actual prediction models. Consequently, the inaccurate parameters will affect the optimality of the DMPC strategy. This paper considers two possible uncertainty sources for nominal prediction models for day-ahead shipboard power system energy management as follows.

1) Uncertain parameters in propulsion load prediction models: A widely-used propulsion load prediction model for vessels is as follows [46]–[49]:

$$P^{\text{pro}}_{i,k} = 2\pi\eta_i K^{\text{Q}}_i \cdot (n^{\text{pro}}_{i,k})^3 |n^{\text{pro}}_{i,k}|,\, k \in \mathcal{K}_t \tag{1}$$

where $P^{\text{pro}}_{i,k}$ is the predicted propulsion load of the propeller in Zone i at time step k using the nominal prediction model, η_i is the comprehensive coefficient of environment and propeller parameters of propeller i, K^{Q}_i is the estimated advance coefficient for obtaining the propulsion load of propeller i, and $n^{\text{pro}}_{i,k}$ is the revolution speed of the propeller in Zone i. When the reference of voyage speeds in one prediction horizon, which can be obtained by the voyage plan, is given, the vessels can control the propeller revolution speeds $n^{\text{pro}}_{i,k}$ to keep the vessel sailing at the reference voyage speeds or routine. Accordingly, the propulsion loads $P^{\text{pro}}_{i,k}$ in one prediction horizon can be calculated according to (1). However, since the propulsion systems of vessels are quite complicated, the coefficient K^{Q}_i is difficult to estimate accurately [46], [48]. Thus, parameter K^{Q}_i is usually biased from $\tilde{K}^{\text{Q}}_i$ of the actual propulsion load prediction model. In comparison, the actual propulsion load prediction model is as follows:

$$\tilde{P}^{\text{pro}}_{i,k} = 2\pi\eta_i \tilde{K}^{\text{Q}}_i \cdot (n^{\text{pro}}_{i,k})^3 |n^{\text{pro}}_{i,k}|,\, k \in \mathcal{K}_t \tag{2}$$

where $\tilde{K}^{\text{Q}}_i$ is the actual advance coefficient for obtaining the propulsion load of the propeller in Zone i and $\tilde{P}^{\text{pro}}_{i,k}$ is the actual propulsion load in Zone i at time step k.

2) Uncertain parameters in generation power prediction models of solar panels: In literature, an autoregressive integrated moving average (ARIMA) model is widely used to predict the generation powers of solar panels [50]–[52]. The form of ARIMA(p,d,q) can be expressed as follows:

$$\phi_i(B)(1-B)^{\text{d}} P^{\text{sol}}_{i,k} = \theta_i(B)\epsilon_k,\, k \in \mathcal{K}_t \tag{3}$$

where $P^{\text{sol}}_{i,k}$ is the generation power of the solar panel in Zone i at time step k using the nominal prediction model, $\phi_i(B) = 1 - \phi_{i,1}B - \phi_{i,2}B^2 - \cdots - \phi_{i,\text{p}}B^{\text{p}}$ and $\theta_i(B) = 1 + \theta_{i,1}B + \theta_{i,2}B^2 + \cdots + \theta_{i,\text{q}}B^{\text{q}}$ represent autoregressive process and moving average process, respectively. Furthermore, B is the backward shift operator, such that $BP^{\text{sol}}_{i,k} = P^{\text{sol}}_{i,k-1}$, and ϵ_k is a stochastic value yield a normal distribution. In the ARIMA model for generation power predictions of solar panels, $\phi_{i,1}, \cdots, \phi_{i,\text{p}}$ and $\theta_{i,1}, \cdots, \theta_{i,\text{q}}$ are parameters that may bias from the parameters of the actual prediction model for generation powers of solar panels in Zone i. The actual prediction model for the generation powers of the solar panels in Zone i is:

$$\tilde{\phi}_i(B)(1-B)^{\text{d}} \tilde{P}^{\text{sol}}_{i,k} = \tilde{\theta}_i(B)\epsilon_k,\, k \in \mathcal{K}_t \tag{4}$$

where $\tilde{P}^{\text{sol}}_{i,k}$ is the actual generation power of the solar panel in Zone i at time step k, $\tilde{\phi}_i$ is the actual parameters of the autoregressive process of the solar panel in Zone i, $\tilde{\theta}_i$ is the actual parameters of the moving process of the solar panel in Zone i.

C. Formulation of the DMPC day-ahead energy management optimization problems

In this subsection, the DMPC day-ahead energy management optimization problems of the shipboard power system will be formulated. The power flows of the optimization problems are labeled in Fig. 2. The model of the batch of batteries in Zone i is as follows:

$$\begin{aligned} \underline{P}_i^{\text{bat}} \leq P_{i,k}^{\text{bat}} \leq \overline{P}_i^{\text{bat}},\, i \in \mathcal{I},\, k \in \mathcal{K}_t \\ \underline{S}_i^{\text{bat}} \leq S_{i,k}^{\text{bat}} \leq \overline{S}_i^{\text{bat}},\, i \in \mathcal{I},\, k \in \mathcal{K}_t \\ S_{i,k-1}^{\text{bat}} + P_{i,k}^{\text{bat}} = S_{i,k}^{\text{bat}},\, i \in \mathcal{I},\, k \in \mathcal{K}_t - \{1\} \end{aligned} \tag{5}$$

where $P_{i,k}^{\text{bat}}$ is the charging power of the batch of batteries in Zone i at time step k, $S_{i,k}^{\text{bat}}$ is the state of charge (SOC) of the batch of batteries in Zone i at time step k, $\overline{P}_i^{\text{bat}}$ and $\underline{P}_i^{\text{bat}}$ are the maximum and minimum charging powers of the battery in Zone i, respectively, $\underline{S}_i^{\text{bat}}$ and $\overline{S}_i^{\text{bat}}$ are the lower and upper SOC boundaries of the batch of batteries in Zone i, respectively. Constraints in (5) represent the charging power bounds, the SOC boundaries, and the SOC accumulation, respectively. The model of DGs is as follows:

$$\begin{aligned} \underline{P}_i^{\text{gen}} \leq P_{i,k}^{\text{gen}} \leq \overline{P}_i^{\text{gen}},\, i \in \mathcal{I},\, k \in \mathcal{K}_t \\ \underline{R}_i^{\text{gen}} \leq P_{i,k}^{\text{gen}} - P_{i,k-1}^{\text{gen}} \leq \overline{R}_i^{\text{gen}},\, i \in \mathcal{I},\, k \in \mathcal{K}_t - \{1\} \end{aligned} \tag{6}$$

where $P_{i,k}^{\text{gen}}$ is the generation power of the diesel generator in Zone i at time step k, $\overline{P}_i^{\text{gen}}$ and $\underline{P}_i^{\text{gen}}$ are the maximum and minimum generation powers of the diesel generator in Zone i, respectively, $\overline{R}_i^{\text{gen}}$ and $\underline{R}_i^{\text{gen}}$ are the maximum and minimum ramping rates of the diesel generator in Zone i, respectively. Constraints in (6) represent the bounds of generation powers and the ramping rate of diesel generators in Zone i. The power balance of Zone i is as follows:

$$P_{i,k}^{\text{gen}} + P_{i,k}^{\text{sol}} = P_{i,k}^{\text{bat}} + P_{i,k}^{\text{pro}} + P_{i,k}^{\text{ser}} + \sum_{j \in \mathcal{J}_i} P_{i,j,k}^{\text{int}},\, i \in \mathcal{I},\, k \in \mathcal{K}_t \tag{7}$$

where $P_{i,k}^{\text{ser}}$ is the service load of Zone i at time step k, and $P_{i,j,k}^{\text{int}}$ is the interconnecting power flow from Zone i to Zone j at time step k. In (7), the generation powers of solar panel $P_{i,k}^{\text{sol}}$ and the propulsion load $P_{i,k}^{\text{pro}}$ are predicted by the nominal prediction models (1) and (3). The constraints of the power exchange among zones can be expressed such that:

$$P_{i,j,k}^{\text{int}} = -P_{j,i,k}^{\text{int}},\, \forall i \in \mathcal{I},\, \forall j \in \mathcal{J}_i,\, \forall k \in \mathcal{K}_t \tag{8}$$

Afterwards, the local objective function J_i of Zone i is as follows:

$$J_i = \sum_{k \in \mathcal{K}_t} \alpha_i (P_{i,k}^{\text{gen}})^2 + \beta_i P_{i,k}^{\text{gen}} \tag{9}$$

where α_i and β_i are the square and linear cost coefficients of the diesel generator in Zone i, respectively. Constraint (9) means that the total cost of generation power of the DG in Zone i during a prediction horizon should be minimized.

Equations (5)-(9) compose the local DMPC energy management optimization problem of Zone i. Accordingly, the global DMPC energy management optimization problem can be expressed as follows:

$$\begin{aligned} \min \sum_{i \in \mathcal{I}} J_i \\ \text{s.t. } (5) - (9),\, \forall i \in \mathcal{I} \end{aligned} \tag{10}$$

where the optimization problem (10) can be solved by distributed algorithms, e.g., the synchronous alternating direction method of multipliers (ADMM) [53], which will be explained in Section II.D. The variables of the global DMPC optimization problem include $P_{i,k}^{\text{bat}}$, $S_{i,k}^{\text{bat}}$, $P_{i,k}^{\text{gen}}$, $P_{i,j,k}^{\text{int}}$. Among the variables, the control actions are $P_{i,k}^{\text{bat}}$ and $P_{i,k}^{\text{gen}}$. After solving the optimization problem (10), the control actions of the first step are implemented by the agents.

From (7), it can be observed that if the propulsion loads and the generation powers of solar panels are not accurately predicted, the power imbalance will emerge if the baseline control actions are not compensated. Thus, Section III will propose an RL-C-DMPC strategy to reduce the power imbalance caused by inaccurate parameters in the nominal prediction models.

D. Synchronous ADMM and its implementation

Synchronous ADMM is an algorithm that can solve global optimization problems in a distributed manner [47], [54]–[56]. At iteration p, Agent i solves the local optimization problem of Zone i as follows:

$$\begin{aligned} \min_{x_i(p), \tilde{x}_{i,j}(p)} L_i(p) = J_i(p) + \sum_{j \in \mathcal{J}_i} \Big(\lambda_i^{\text{T}}(p)(\tilde{x}_{i,j}(p) - \overline{z}_{i,j}(p)) + \\ \frac{\rho}{2} \left\| \tilde{x}_{i,j}(p) - \overline{z}_{i,j}(p) \right\|_2^2 \Big) \\ \text{s.t. } \big(x_i(p),\, \tilde{x}_{i,j}(p)\big) \in \mathcal{X}_i \end{aligned} \tag{11}$$

where ρ is the penalty parameter of the augmented Lagrangian term, $x_i(p) = [P_{i,1}^{\text{bat}}(p), ..., P_{i,|\mathcal{K}_t|}^{\text{bat}}(p), S_{i,1}^{\text{bat}}(p), ..., S_{i,|\mathcal{K}_t|}^{\text{bat}}(p), P_{i,1}^{\text{gen}}(p), ..., P_{i,|\mathcal{K}_t|}^{\text{gen}}(p)]^{\text{T}}$ is the vector of local variables of Zone i in one prediction horizon at iteration p, $\tilde{x}_{i,j}(p) = [P_{i,j,1}^{\text{int}}(p), ..., P_{i,j,|\mathcal{K}_t|}^{\text{int}}(p)]^{\text{T}}$ is the vector of interconnecting power flows of Zone i in one prediction horizon at iteration p, and $\mathcal{X}_i$ is the set of power system constraints of Zone i, such that:

$$\begin{aligned} \mathcal{X}_i = \{(x_i(p), \tilde{x}_{i,j}(p) | (x_i(p), \tilde{x}_{i,j}(p)) \text{ satisfy } (5) - (7), \\ \text{and } (9)\} \end{aligned} \tag{12}$$

Furthermore, in (11), $\overline{z}_{i,j}(p)$ is calculated by:

$$\overline{z}_{i,j}(p) \leftarrow \frac{1}{2}(\tilde{x}_{i,j}(p-1) + z_{i,j}(p-1)) \tag{13}$$

where $z_{i,j}(p)$ is the local copy of $\tilde{x}_{j,i}(p)$ that is received from Zone j via the communication network at iteration p. Fig. 4 illustrates a physical shipboard power system and its communication architecture of synchronous ADMM. In Fig. 4, each agent is equipped with a GPS to ensure the same clock.

The processes of the synchronous ADMM algorithm are shown in **Algorithm 1**. Two stopping criteria are concerned:

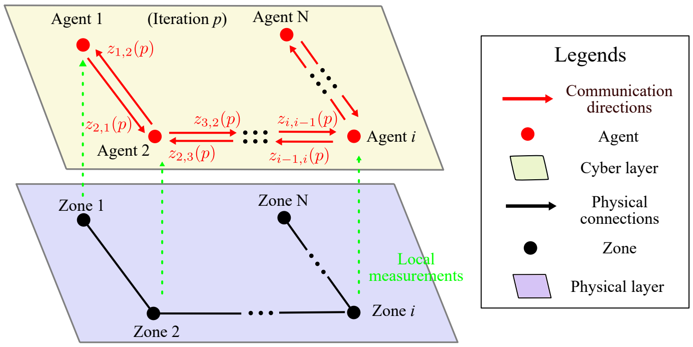


Fig. 4. The communication architecture of synchronous ADMM

```
Algorithm 1: Synchronous ADMM
1: Set the tolerance gaps $\xi_1$, $\xi_2$, and the penalty parameter $\rho$.
2: At the beginning of the time step $t$, the agents collect parameters of the local optimization problem (11) from local measurements. The agents initialize Lagrangian multiplier $\lambda_i(0)$ and set $p \leftarrow 1$.
3: while $p \leq$NP or (14) is not satisfied do
4:     For each Zone $i$, where $i \in \mathcal{I}$, Agent $i$:
5:     Calculates $\overline{z}_{i,j}(p)$ via (13). Updates
       $\lambda_i(p) \leftarrow \lambda_i(p-1) + \rho \cdot \left(\tilde{x}_{i,j}(p-1) - \overline{z}_{i,j}(p-1)\right)$.
6:     Solves local optimization problem (11) and obtains
       $x_i(p)$ and $\tilde{x}_{i,j}(p)$. Duplicates variables
       $z_{j,i}(p) = -\tilde{x}_{i,j}(p)$. Waits $t_{\mathrm{sol}}$.
7:     Sends $z_{j,i}(p)$ to all neighbor Zones $j$, and receives
       $z_{i,j}(p)$ from neighbor Zones $j$, where $j \in \mathcal{J}_i$. Waits
       $t_{\mathrm{tr}}$.
8:     $p \leftarrow p + 1$
9: end while
```

the number of iterations exceeds the maximum number of iterations NP, or tolerance gaps (14) are satisfied.

$$\|\lambda_i(p) - \lambda_i(p-1)\|_2^2 \leq \xi_1,\ \|z_i(p) - z_i(p-1)\|_2^2 \leq \xi_2 \quad (14)$$

In **Algorithm 1**, t_{sol} is the set solution time for each agent to solve its local optimization problem (11), and t_{tr} is the set communication time for sending information among agents. An illustrative timeline in one iteration of synchronous ADMM for a communication network with three agents (each agent connects to another two) is shown in Fig. 5. In Fig. 5, at the beginning of one iteration, each agent implements Lines 5 and 6. The solution time for each agent may be different, as described in the red bars in Fig. 5. After waiting t_{sol}, the agents send information to other agents. The communication times for each agent to receive all the required information may differ, as illustrated in green bars in Fig. 5. After waiting t_{tr}, a new iteration starts.

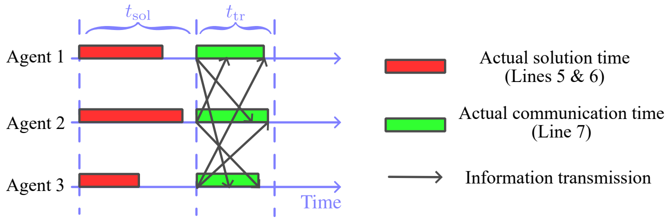


Fig. 5. The timeline in one iteration of synchronous ADMM: A three-agent example

III. The proposed RL-C-DMPC strategy

A. Control structure

The control structure of the proposed RL-C-DMPC for Zone i is shown in Fig. 6. Three main modules are included, i.e., the DMPC module, the RL module, and the actual shipboard power system of Zone i. The DMPC module determines the control actions of one prediction horizon, and the control actions of the first step are marked as $\hat{P}^{\mathrm{gen}}_{i,t}$ and $\hat{P}^{\mathrm{bat}}_{i,t}$. Then $\hat{P}^{\mathrm{gen}}_{i,t}$ is compensated by multiplying the sum of one and the compensation power rate, which is the output of the RL module.

In Fig. 6, the states for prediction $v^{\mathrm{P}}_{i,t} = [n^{\mathrm{pro}}_{i,t}, P^{\mathrm{sol}}_{i,t-\mathrm{p\cdot d}}, ..., P^{\mathrm{sol}}_{i,t-1}]$, where $k \in \mathcal{K}_t$, include the revolution speed of the propeller in one prediction horizon starting from the current time step t and the historical generation powers of the solar panels. After a truncated section in Fig. 6, the states for prediction are truncated to the states (inputs) of the DQN, i.e., $w^{\mathrm{P}}_{i,t} = [n^{\mathrm{pro}}_{i,t}, P^{\mathrm{sol}}_{i,t-\mathrm{p\cdot d}}, ..., P^{\mathrm{sol}}_{i,t-1}]$. Furthermore, because the ramping rate of DGs should be considered in DG generation power compensation, the DG generation power at the last time step should also be included in states (inputs) of the DQN, i.e., $w^{\mathrm{O}}_{i,t} = [P^{\mathrm{gen}}_{i,t-1}]$.

The actual system state $v^{\mathrm{O}}_{i,t} = [S^{\mathrm{bat}}_{i,t-1}, P^{\mathrm{gen}}_{i,t-1}]$ includes the SOC level of the batch of batteries and the DG generation power at the last time step $t-1$ in Zone i. According to the states for prediction $v^{\mathrm{P}}_{i,t}$ and the actual system state $v^{\mathrm{O}}_{i,t}$, the local optimization problem (11) can be formulated by nominal prediction models (1) and (3), and nominal system models (5)-(7) and (9). Afterward, the agents of zones solve the local optimization problem (11) distributively via synchronous ADMM and the communication architecture in Fig. 4. Then, the control actions of zones can be obtained, i.e., $\hat{P}^{\mathrm{gen}}_{i,t}$ and $\hat{P}^{\mathrm{bat}}_{i,t}$.

By multiplying $\hat{P}^{\mathrm{gen}}_{i,t}$ and $(1+u^{\mathrm{F}}_{i,t})$, the compensated generation power of diesel generator of Zone i at time step t can be implemented to the actual shipboard power system of Zone i. Afterward, the new time step $t+1$ begins.

B. Implementation of the proposed RL-C-DMPC

In Fig. 6, Agent i mainly includes a DMPC module and an RL module. To train the RL modules of the agents, this paper adopts a centralized training mechanism and a distributed testing (implementation) mechanism, as shown in Fig. 7. Note that the details of the training algorithm, i.e., VDN, will be illustrated in Section III.C.

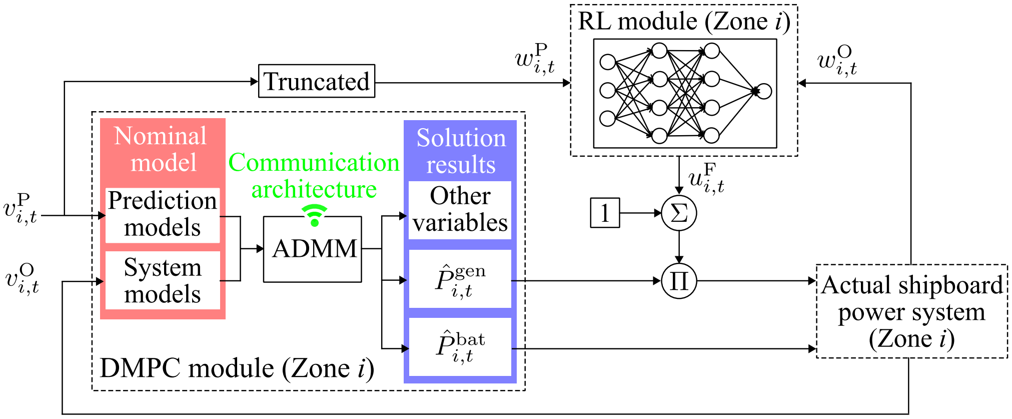
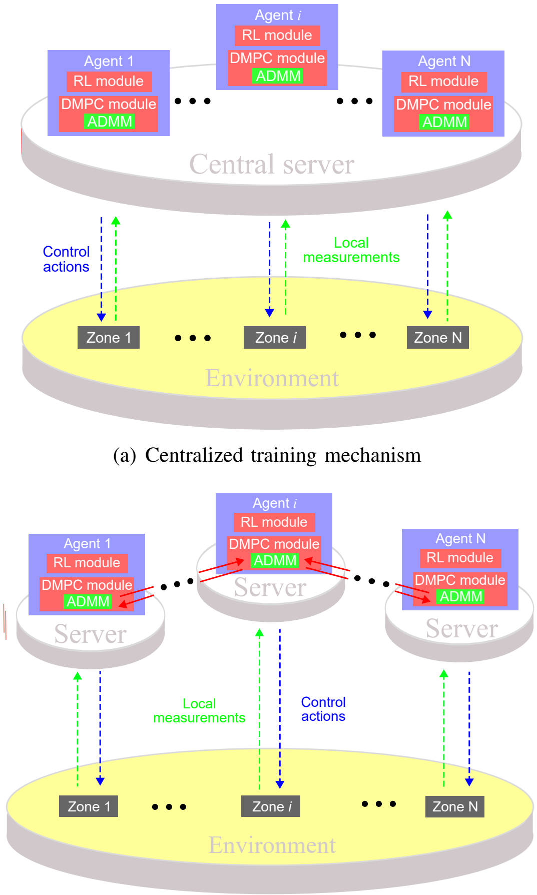


Fig. 6. The control structure of Agent i of the proposed RL-C-DMPC.



(a) Centralized training mechanism

(b) Distributed testing mechanism

Fig. 7. Centralized training and distributed testing mechanisms.

In Fig. 7(a), the central server collects the local measurements obtained by the sensors in each zone. The agents whose control structures are shown in Fig. 6 determine the compensated control actions according to the local measurements. Then, the control actions output by the agents are implemented in the zones in the environment to obtain rewards, the next states, and the signal of whether the episode is over. Afterward, the “Global loss” of VDN is calculated and distributed to the policy network in the RL module of each agent via the central server. Accordingly, the networks of the agents can be updated and trained. Fig. 7(b) illustrates how the agents are distributively tested. In Fig. 7(b), each agent collects the local measurements and implements the synchronous ADMM as explained in Section II.D. Afterward, the agents output the compensated control actions, and the compensated control actions, i.e., compensated generation power of diesel generator, are implemented in each zone.

The authors adopt a distributed testing (implementation) manner because distributed control frameworks can enhance communication robustness compared to centralized control frameworks [57], [58]. In distributed communication architectures, when failures, e.g., missing data and delays, occur, the influences of the failures on the performance of the energy management strategies can be reduced. To evaluate the communication robustness of RL-C-DMPC, in Section IV.C, we will compare the optimality of RL-C-DMPC and RL-compensated MPC (RL-C-MPC), which has a centralized control framework when failures occur. Two failure scenarios will be studied in the evaluation of communication robustness in Section IV.C. First, failures occur when the local measurements are collected by the sensors and sent to the agents. Second, failures occur when the control actions output by the agents are sent to the zones for implementation. Further details of the comparison will be explained in Section IV.C.

C. VDN-based centralized training mechanism

To train the DQN in the RL module shown in Fig. 6, this paper proposes a centralized training mechanism based on VDN as illustrated in Fig. 8. The basic idea of VDN is to obtain the global Q value by summing up the local Q values. Then the global losses derived from the global Q value are applied to train local agents.

In Fig. 8, index i of State_i, Next_state_i, Q_eval_i, Action_i, Q_target_i, and y_i represents Zone i. Furthermore, State_i, Next_state_i, Q_eval_i, Action_i, and Q_target_i are state, next state after implementing the action, the Q value of policy network, action, and Q value of target network of Zone i, respectively. State_i in Fig. 8 corresponds to $w^{\mathrm{P}}_{i,t}$ and

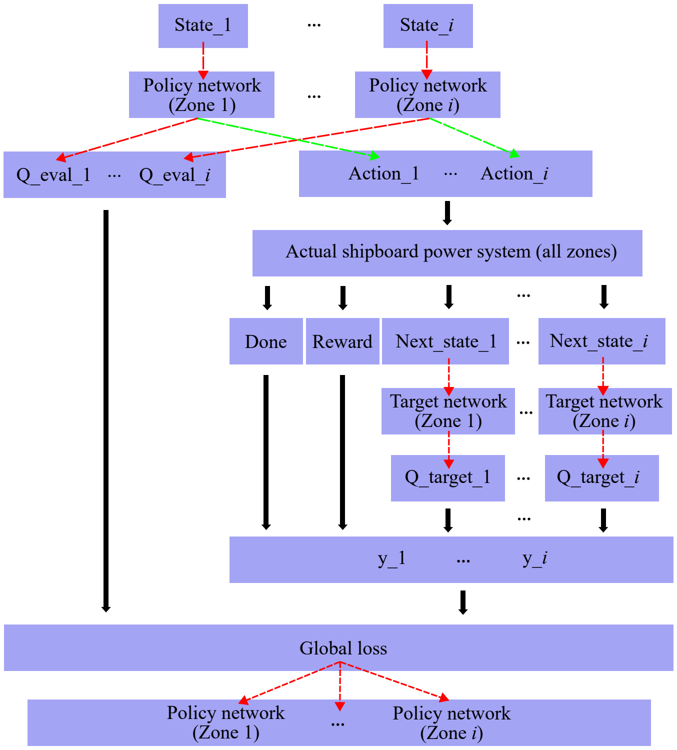


Fig. 8. Value-decomposition-network-based centralized training mechanism of DQNs in the proposed RL-C-DMPC.

$w_{i,t}^{\mathrm{O}}$ mentioned in the control structure of Fig. 6. Additionally, "Reward" and "Done" in Fig. 8 are the global reward obtained after the action is implemented, and the signal of whether the episode is over, respectively. If the episode is over, "Done" is set to 1, and if not, "Done" is set to 0. Furthermore, y_i can be obtained by:

$$y_i = \text{Reward} + \gamma \cdot \text{Q\_target\_}i \cdot (1 - \text{Done}) \tag{15}$$

where γ is a discount factor. Moreover, the "Global loss" is obtained by:

$$\text{Global loss} = \sum_{i \in \mathcal{I}} (y_i - \text{Q\_eval\_}i)^2 \tag{16}$$

In Fig. 8, after the value and gradients of "Global loss" are distributed to the policy networks of all zones, the policy networks of all zones are updated distributively. Then, the current training episode will be over. This design adopts the backpropagation to update the weights of DQNs and the adaptive moment estimation algorithm as the optimizer.

After the centralized training process, the trained policy networks of zones are stored. The stored networks can be applied for distributed testing (or implementation). The testing mechanism is illustrated in Fig. 9. The following Section III.D will introduce the design of the "Actual shipboard power system (for all zones)" in Fig. 8 and Fig. 9, which is the environment of the training and testing mechanisms.

D. Design of the actual shipboard power system environment

Compared to the nominal models, the actual shipboard power system contains the actual prediction models with actual parameters. The environment in training and testing

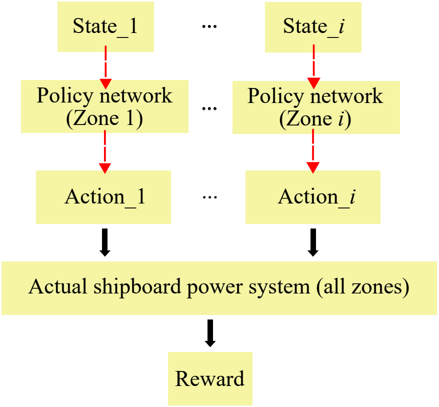


Fig. 9. Distributed testing mechanism of DQNs in the proposed RL-C-DMPC.

outputs the value of the next state (Next_state_i), reward ("Reward"), and the signal of whether the episode is over ("Done") according to the input actions of all zones. This subsection illustrates how the outputs are derived from the inputs.

Because the input actions of DQNs are discrete, we design that one action value of agent i corresponds to one compensation power rate, such that:

$$P_{i,t}^{\mathrm{cp}} = \underline{P}_{i,t}^{\mathrm{cp}} + \text{Action\_}i \cdot \Delta P_{i,t}^{\mathrm{cp}} \tag{17}$$

where $\underline{P}_{i,t}^{\mathrm{cp}}$ is the starting value of compensation power rate and $\Delta P_{i,t}^{\mathrm{cp}}$ is the ratio of linear increment. Then, the compensated DG generation powers can be obtained such that:

$$\tilde{P}^{\text{gen}}_{i,t} = \max\{\min\{\hat{P}^{\text{gen}}_{i,t} \cdot (1 + P^{\text{cp}}_{i,t}), \overline{P}^{\text{gen}}_{i,t}\}, \underline{P}^{\text{gen}}_{i,t}\} \tag{18}$$

where $\tilde{P}^{\text{gen}}_{i,t}$ is the compensated DG generation power. In (17), the final compensated DG generation power is bounded by the maximum and minimum generation powers. Then the actual propulsion load and the actual generation power of solar panel in Zone i, i.e., $\tilde{P}^{\text{pro}}_{i,t}$ and $\tilde{P}^{\text{sol}}_{i,t}$, can be obtained by (2) and (4), respectively.

Afterwards, we design the global reward ("Reward") for training and testing as the negative value of the square of power imbalance of the shipboard power system, given as:

$$\begin{aligned}\text{Reward} = -(\sum_{i\in\mathcal{I}} \tilde{P}^{\text{gen}}_{i,t} + \tilde{P}^{\text{sol}}_{i,t} - \hat{P}^{\text{bat}}_{i,t} - \tilde{P}^{\text{pro}}_{i,t} - P^{\text{ser}}_{i,t})^2 \\ -\sigma \cdot \text{Penalty}\end{aligned} \tag{19}$$

where $\hat{P}^{\text{bat}}_{i,t}$ represents the charging power of the batch of batteries in Zone i at time step t obtained by solving the DMPC energy management optimization problem as shown in Fig. 6, σ is the signal of whether the ramping rate boundaries are satisfied. In detail, if $\tilde{P}^{\text{gen}}_{i,t} - P^{\text{gen}}_{i,t-1} \geq \overline{R}^{\text{gen}}_i$ or $\tilde{P}^{\text{gen}}_{i,t} - P^{\text{gen}}_{i,t-1} \leq \underline{R}^{\text{gen}}_i$, we have $\sigma = 1$, otherwise $\sigma = 0$. In (19), "Penalty" is a sufficiently large positive value.

Because of the receding horizon mechanism of DMPC, the RL-C-DMPC proposed in this paper only compensates for DG generation powers at the current time step t. Thus, one training episode includes only one single time step, so "Done" is set to 1.

Fig. 10. Solar panels in the lab of Lingshui Port in the area of the voyage.

Table I: Parameters of the diesel generators and batches of batteries in the case study

DG	$\underline{P}^{\text{gen}}_i$	$\overline{P}^{\text{gen}}_i$	$\underline{R}^{\text{gen}}_i$	$\overline{R}^{\text{gen}}_i$
1-3	0 kW	2000 kW	-1600 kW	1600 kW
Battery	$\underline{P}^{\text{bat}}_i$	$\overline{P}^{\text{bat}}_i$	$\underline{E}^{\text{bat}}_i$	$\overline{E}^{\text{bat}}_i$
1	-375 kW	375 kW	37.5 kW.h	337.5 kWh
2	-375 kW	375 kW	37.5 kW.h	337.5 kWh
3	-525 kW	525 kW	52.5 kW.h	472.5 kWh

E. A method for selecting the range of compensation power rates

DQNs can output actions according to the states. Selecting a small range of compensation rates may result in insufficient compensation. On the other hand, selecting a large range of compensation rates with a small-scale action space may lead to large compensation errors. Furthermore, selecting a large range of compensation rates with a large-scale action space may take a lot of time to train DQNs. Thus, selecting a proper range of compensation rates is important for RL-C-DMPC. Accordingly, this subsection proposes a specific method to select proper ranges of compensation rates for RL-C-DMPC for shipboard power system energy management.

If the boundaries of power imbalance can be obtained, the boundaries of compensation power rates can be estimated. For example, if we know the boundaries of uncertain parameters K^{Q}_i, ϕ, and θ, we can calculate the largest and smallest gaps between the proportion loads and generation powers of solar panels using the nominal and actual prediction models. These gaps are marked as $DP^{\text{pro,max}}_i$, $DP^{\text{pro,min}}_i$, $DP^{\text{sol,max}}_i$, and $DP^{\text{sol,min}}_i$. Then the lower and upper boundaries of power imbalances can be obtained by $DP^{\text{pro,min}}_i - DP^{\text{sol,max}}_i$ and $DP^{\text{pro,max}}_i - DP^{\text{sol,min}}_i$. Afterward, the range of compensation rates can be obtained as follows:

$$\begin{aligned}\overline{Range} &= \zeta \cdot \sum_{i\in\mathcal{I}} (DP^{\text{pro,max}}_i - DP^{\text{sol,min}}_i) / \sum_{i\in\mathcal{I}} \overline{P}^{\text{gen}}_i \\ \underline{Range} &= \zeta \cdot \sum_{i\in\mathcal{I}} (DP^{\text{pro,min}}_i - DP^{\text{sol,max}}_i) / \sum_{i\in\mathcal{I}} \overline{P}^{\text{gen}}_i\end{aligned} \tag{20}$$

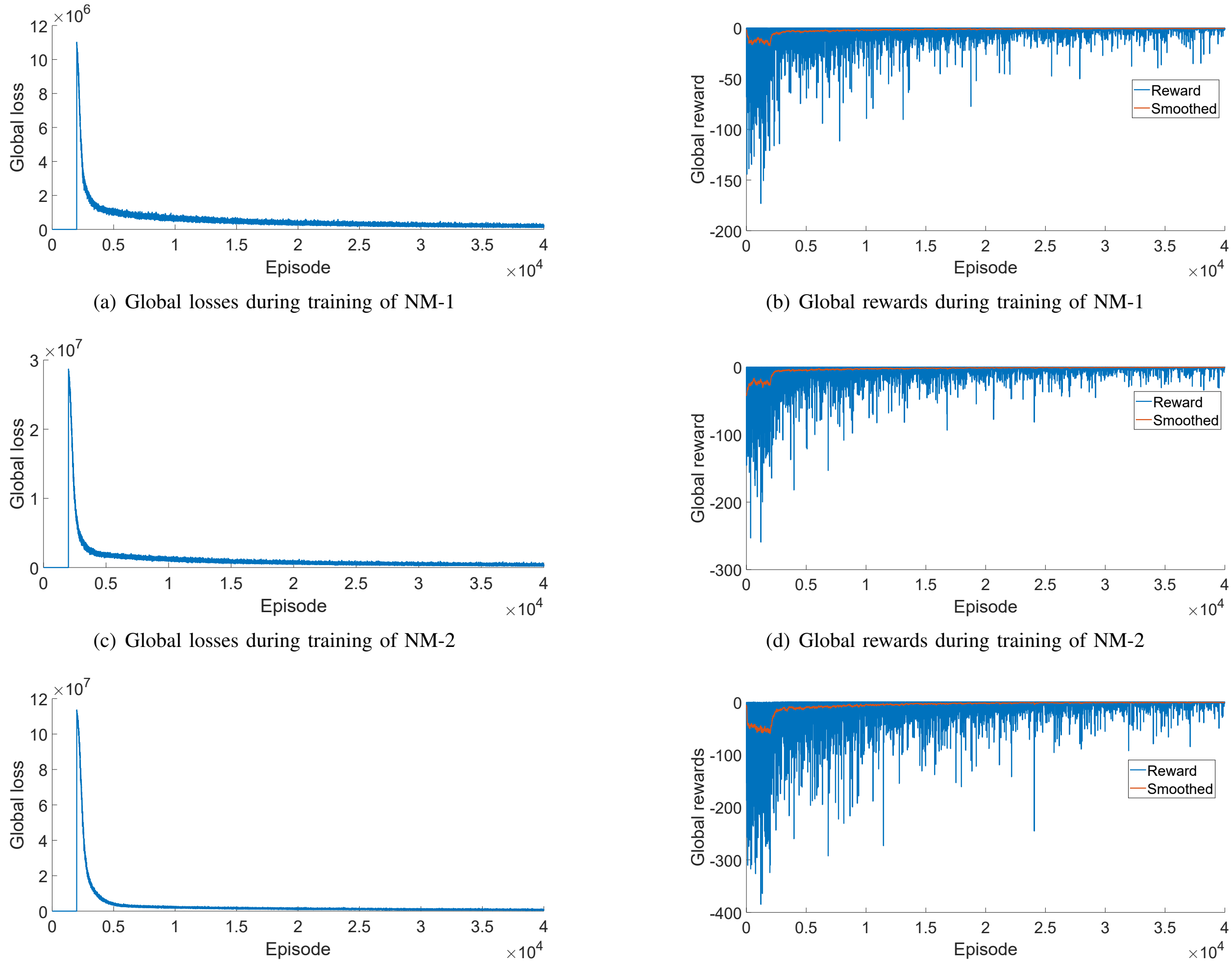
where ζ is a margin coefficient larger than 1 to avoid insufficient compensation, e.g., 1.2. Furthermore, $\overline{Range}$ and $\underline{Range}$ are the upper and lower boundaries of compensation rates, respectively.

IV. Case Study

A. Basic settings

This section tests the proposed RL-C-DMPC strategy for the energy management of the shipboard power system in Fig. 2. The lengths of a time step and a prediction horizon are 1 h and 24 h, respectively. The parameters of the DGs and the batches of batteries are shown in Table I. The rated powers of the Service load 1 and 2 are 400 kW and 200 kW, respectively. The rated Propulsion loads 1 and 2 are both 2200 kW. Furthermore, since the rated power of solar panels is around 1 kW/m^2, according to the spaces of "Yukun", the rated generation powers of Solar panels 1 to 3 are 30 kW, 90 kW, and 30 kW, respectively.

To show the effectiveness of the proposed RL-C-DMPC on handling uncertainties, the power imbalances with DMPC (no compensation) and those with RL-C-DMPC are compared. The DMPC solves the optimization problem (10) using nominal prediction models via synchronous ADMM, and the RL-C-DMPC treats the solution of DMPC as the baseline control. During centralized training, the optimization problem (10) is solved by "quadprog" function in Matlab. During distributed testing, (10) is solved by ADMM where the local optimization problems are solved by "quadprog" function in Matlab. The DQNs of RL-C-DMPC are trained and tested via



(a) Global losses during training of NM-1

(b) Global rewards during training of NM-1

(c) Global losses during training of NM-2

(d) Global rewards during training of NM-2

(e) Global losses during training of NM-3

(f) Global rewards during training of NM-3

Fig. 11. Global losses and rewards during training processes of DQNs in the proposed RL-C-DMPC.

the mechanisms in Fig. 8 and Fig. 9 under the environment mentioned in Section III.C.

The revolution speeds of the propeller are collected by the measurement system on "YuKun" during its sailing on the Yellow Sea in the northeast of China. The historical solar generation powers are collected by the solar panels in the lab of the Lingshui Port near Yellow Sea (the same area as the voyage), as shown in Fig. 10. Because large p, d, and q values of ARIMA(p,d,q) model increase the computational burden, this paper adopts ARIMA models whose p+d+q values are no larger than five. Since ARIMA(2,1,1) most fits the solar radiation data in Lingshui Port among ARIMA models satisfying p+d+q≤5, ARIMA(2,1,1) is adopted in this paper. The parameters in nominal prediction models biased from those in actual prediction models are listed in Table II, where else three cases with different nominal models are used in DMPC and considered for testing the general performance of the proposed RL-C-DMPC. Furthermore, "AM" represents the parameters of the actual prediction model, and "NM-1" to "NM-3" represent the parameters of the nominal prediction models in the three cases.

Table II: The biased parameters between nominal and actual prediction models (AM: actual model, NM: nominal model)

Parameters	AM	NM-1	NM-2	NM-3
K_1^{Q}	1	0.98	1	0.98
K_2^{Q}	1	1	0.97	0.97
ϕ_2	1	0.97	0.97	0.97

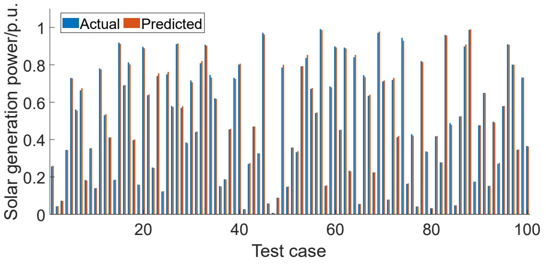


Fig. 12. Actual and predicted solar generation powers for 100 test cases.

Furthermore, we collect 5000 groups of training data and 100 groups of testing data (i.e., 100 test cases). The predicted solar generation powers versus the actual solar generation powers for 100 test cases are shown in Fig. 12 (in p.u. values). The batch size of the training process is 1000. For every 40 episodes, the target network is updated according to the current policy network. The dimension of the hidden layer is 128 and the rectified linear unit is selected as the activation function. The training and testing processes are implemented on Python 3.11.2. The maximum number of training episodes for all cases is 40000.

Moreover, to evaluate the frequency stability when power imbalances occur, the following frequency regulation model is used [59]:

$$\sum_{i\in\mathcal{I}} \tilde{P}^{\text{gen}}_{i,t} + \tilde{P}^{\text{sol}}_{i,t} - \hat{P}^{\text{bat}}_{i,t} - \tilde{P}^{\text{pro}}_{i,t} - P^{\text{ser}}_{i,t} = \sum_{i\in\mathcal{I}} (\text{KG}^{\text{gen}}_{i} + \text{KG}^{\text{pro}}_{i}) \cdot (f - f_0) \tag{21}$$

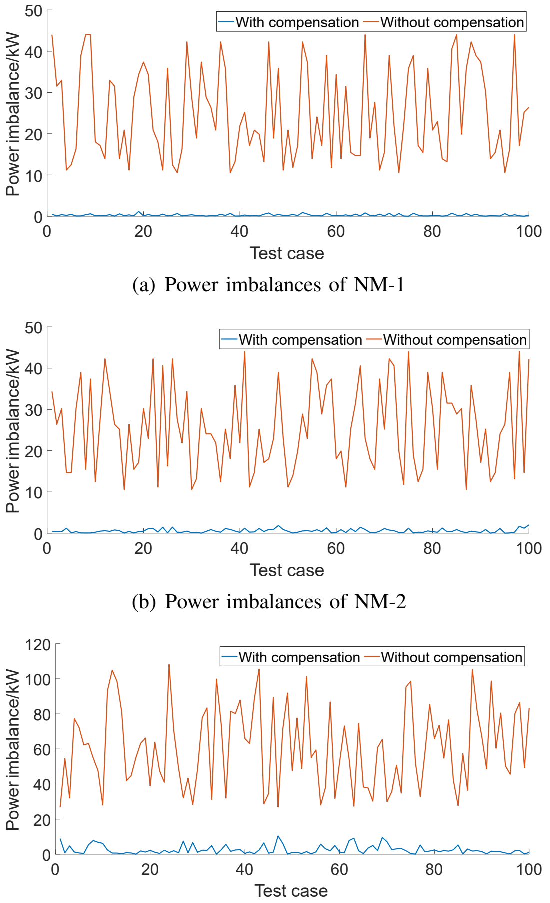
where $\text{KG}^{\text{gen}}_{i}$ is the damping coefficient of the DG in zone i, $\text{KG}^{\text{pro}}_{i}$ is the damping coefficient of the propulsion load in zone i, f is the frequency of the shipboard power system, and $f_0 = 50$ is the fundamental frequency. The left-hand side of (21) is the power imbalance, and the right-hand side of (21) is the sum of damping coefficients multiplied by the frequency variation. The damping coefficients of DG1 to DG3 are 2, and the damping coefficients of two propulsion loads are 2.2 [59]–[61]. This simulation assumes that only the DGs and the propulsion loads join in the frequency regulation.

B. Training and testing results of NM-1 to NM-3

The training processes of NM-1 to NM-3 are shown in Fig. 11. From Fig. 11, it can be observed that for NM-1 to NM-3, the global losses decrease as the training processes proceed. The global losses converge before 40000 episodes are trained. Furthermore, the global rewards as mentioned in (19) increase and approach zero as the training processes proceed. Since for NM-3, the deviations of parameters are larger than NM-1 and NM-2, the global losses and rewards are the largest among NM-1 to NM-3 during training. The global losses and rewards curves show the convergence and effectiveness of the proposed RL-C-DMPC.

Fig. 13 and Fig. 14 show the testing results of NM-1 to NM-3. Fig. 13 shows the power imbalances of 100 test cases of NM-1 to NM-3. The red and blue curves show the power imbalances when implementing the baseline control without compensation and the baseline control with compensation to the actual shipboard power system environment, respectively. Since the deviations of the parameters of NM-3 is larger than NM-1 and NM-2, the power imbalance is the largest among NM-1 to NM-3 on average for implementing the baseline control without compensation. From Fig. 13, it can be observed that, with compensation, i.e., our proposed RL-C-DMPC, the power imbalances can be largely reduced for NM-1 to NM-3. The reduction ratio can be around 90% of the power imbalance of the baseline control on average.

Fig. 14 further shows the details of the compensation powers of NM-1 to NM-3 for 100 test cases. Since the compensated generation power is the sum of the generation power obtained by DMPC and the compensation power, negative compensation powers in Fig. 14 mean that, after compensation, the compensated generation powers of the generators are reduced compared to the generation powers obtained by DMPC. From Fig. 14, it can be observed that, after training, all the agents are involved in compensating the power imbalances and contributing to rectifying the power imbalances. In a few test cases among 100 test cases of NM-1 to NM-3, although some agents contribute negatively to enlarge the power imbalance, e.g., in the 4th and 5th test cases of NM-1, the total contributions of the agents can successfully rectify the power imbalance as shown in Fig. 13. Finally, after compensation, the details of the compensated generation powers and the consumed powers for NM-1 to NM-3 are shown in Fig. 15. Thus, from Fig. 13, Fig. 14, and Fig. 15, it can be observed that the agents work distributed and succeed to rectify the power imbalances caused by the inaccurate parameters in nominal prediction models with our proposed RL-C-DMPC.



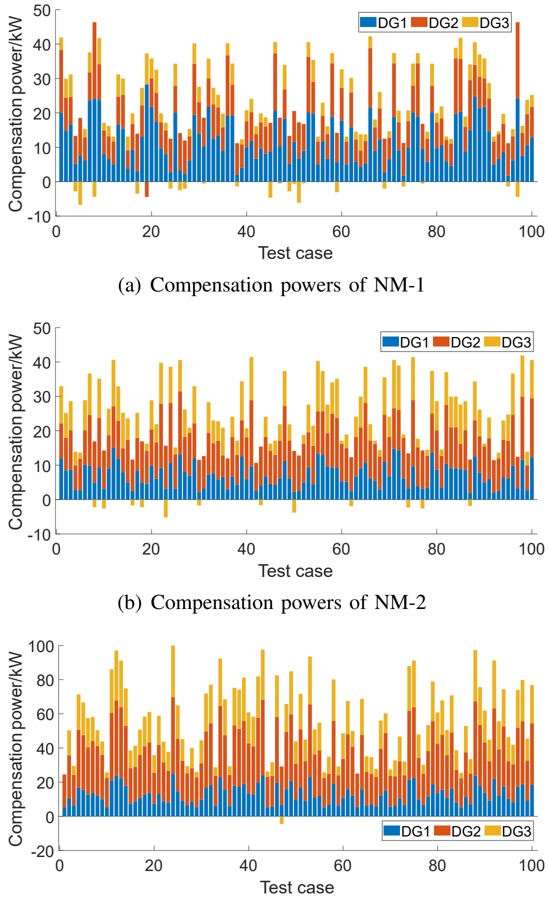
(a) Power imbalances of NM-1

(b) Power imbalances of NM-2

(c) Power imbalances of NM-3

Fig. 13. Power imbalance comparisons between those with and without compensation for 100 different test cases.

Fig. 16 shows the frequency variation simulation results. The frequency variation (in percentage) is defined as $(f - f_0)/f_0$, where f is obtained via model (21). Fig. 16 shows that, without compensation, the frequency variations may exceed the security range (normally from -5% to 5%). However, the frequency variations can remain in the security range with

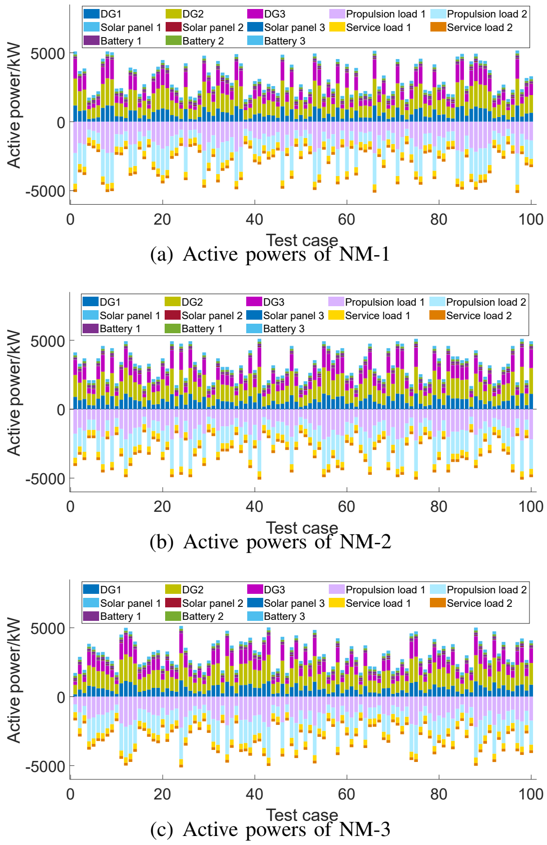


(a) Compensation powers of NM-1

(b) Compensation powers of NM-2

(c) Compensation powers of NM-3

Fig. 14. Compensation powers of agents for NM-1 to NM-3 for 100 different test cases.



(a) Active powers of NM-1

(b) Active powers of NM-2

(c) Active powers of NM-3

Fig. 15. Active power details for NM-1 to NM-3 for 100 different test cases.

compensation. Thus, our proposed RL-C-DMPC is vital for the shipboard power system frequency stability.

C. Communication robustness evaluation of RL-C-DMPC

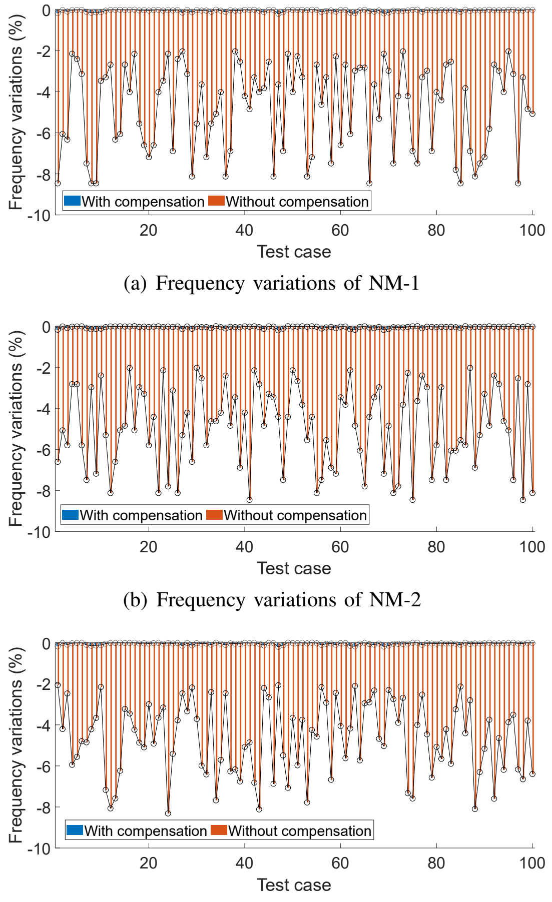
To show the communication robustness of RL-C-DMPC, this subsection will compare the optimality of RL-C-MPC and the proposed RL-C-DMPC under failures. Two scenarios of failures explained in Section III.B are concerned. The control structure and the training/testing mechanism of RL-C-MPC are shown in Fig. 17 and Fig. 18, respectively.

As shown in Fig. 17, there is only one central agent with an MPC module and an RL module to determine control actions. DQN is used in the RL module of RL-C-MPC, as is the case with RL-C-DMPC. The training data of RL-C-MPC is the same as RL-C-DMPC. In Fig. 17, $\mathbf{v}_t^{\mathrm{O}}$, $\mathbf{v}_t^{\mathrm{P}}$, $\hat{\mathbf{P}}_t^{\mathrm{gen}}$, $\hat{\mathbf{P}}_t^{\mathrm{bat}}$, $\mathbf{w}_t^{\mathrm{P}}$, $\mathbf{w}_t^{\mathrm{O}}$, and $\mathbf{u}_t^{\mathrm{F}}$ are the combinations of variables $v_{i,t}^{\mathrm{O}}$, $v_{i,t}^{\mathrm{P}}$, $\hat{P}_{i,t}^{\mathrm{gen}}$, $\hat{P}_{i,t}^{\mathrm{bat}}$, $w_{i,t}^{\mathrm{P}}$, $w_{i,t}^{\mathrm{O}}$ and $u_{i,t}^{\mathrm{F}}$ for all zones, i.e., $i \in \mathcal{I}$, respectively. The algorithm for solving the global optimization problem (10) is the interior point method. Regarding the implementation of RL-C-MPC, as illustrated in Fig. 18, the local measurements are collected and sent to the central server. Then, the central agent, whose control structure is shown in Fig. 17, outputs the compensated control actions. Afterward, the central agent distributes the compensated control actions to all the zones via communication architecture.

In the simulation of communication robustness, we study the performances when the failure rates vary from 10% to 50%. The failure rate is the probability that collected observation data, information, or control actions may fail to be sent to the target agent(s) or zones. For each failure rate value, 30 cases are tested. When failures occur during collecting and sending observation data, we assume that the observation data remains the one received at the last time step [57], [62], [63]. When failures occur during sending control actions, we assume that the control actions remain the ones of the last time step [57], [64], [65].

To evaluate the performances of RL-C-MPC and RL-C-DMPC under failure scenarios, we define the optimality gaps of RL-C-DMPC and RL-C-MPC as $|J^{\mathrm{D}} - J^{*}|/J^{*}$ and $|J^{\mathrm{C}} - J^{*}|/J^{*}$, where J^{D} and J^{C} are $\sum_{i\in\mathcal{I}} J_i$ (global objective function values) obtained by RL-C-DMPC and RL-C-MPC, respectively, and J^{*} is the optimal global objective function value when no communication failures occur. Moreover, we also define the relative gap $|J^{\mathrm{D}} - J^{\mathrm{C}}|/J^{\mathrm{C}}$ to quantify the comparison results between RL-C-DMPC and RL-C-MPC.

The optimality gaps of RL-C-MPC and RL-C-DMPC under failure scenarios are shown in Fig. 19. In Fig. 19, the black crosses represent the mean values of the optimality gaps regarding each failure rate. The mean values ("Mean") and variances ("Var") of the optimality gaps for RL-C-DMPC and



(a) Frequency variations of NM-1

(b) Frequency variations of NM-2

(c) Frequency variations of NM-3

Fig. 16. Frequency simulation results of NM-1 to NM-3.

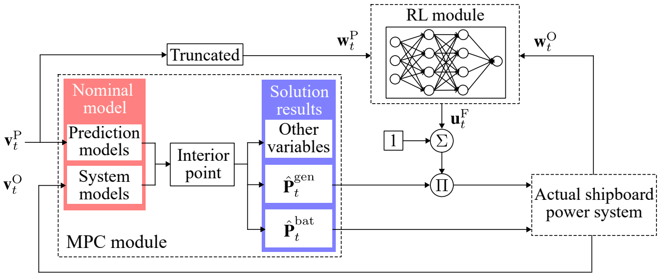

Truncated
$\mathbf{w}_t^{\mathrm{P}}$
RL module
$\mathbf{w}_t^{\mathrm{O}}$
Nominal model
Prediction models
System models
$\mathbf{v}_t^{\mathrm{P}}$
$\mathbf{v}_t^{\mathrm{O}}$
Interior point
Solution results
Other variables
$\hat{\mathbf{P}}_t^{\mathrm{gen}}$
$\hat{\mathbf{P}}_t^{\mathrm{bat}}$
$\mathbf{u}_t^{\mathrm{F}}$
1
Σ
Π
Actual shipboard power system
MPC module

Fig. 17. The control structure of the central agent of RL-C-MPC.

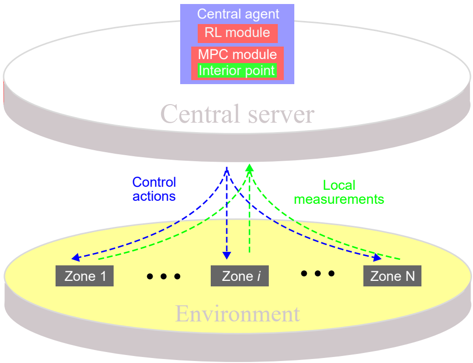


Fig. 18. Training/testing mechanism of RL-C-MPC.

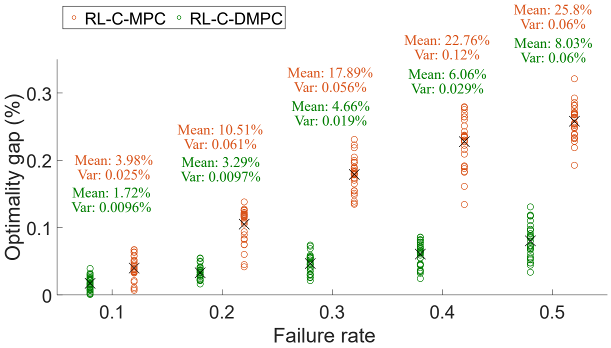


Fig. 19. Optimality gaps of RL-C-MPC and RL-C-DMPC under failure scenarios.

RL-C-MPC are also marked in Fig. 19. From Fig. 19, it can be observed that when failure rates increase, mean values and variances of optimality gaps generally increase. Moreover, for all failure rate values, the variances of optimality gaps of RL-C-DMPC are smaller than those of RL-C-MPC. This demonstrates that the performance of RL-C-DMPC is more stable than RL-C-MPC when communication failures occur. Furthermore, the mean values of optimality gaps of RL-C-DMPC are smaller than those of RL-C-MPC. In details, the relative gaps are 0.5687, 0.6873, 0.7394, 0.7337, and 0.6889, respectively. From the relative gaps, it can be observed that when the failure rate increases, the relative gaps generally increase, so the effectiveness of the communication robustness enhancement of RL-C-DMPC increases. Thus, our proposed distributed control strategy, i.e., RL-C-DMPC, is more robust than the centralized control strategy, i.e., RL-C-MPC, when communication failures occur.

V. Conclusion

This paper proposed a novel RL-C-DMPC strategy to address the issue of power imbalances in shipboard power systems caused by inaccurate predictions resulting from uncertain parameters in the nominal prediction models of DMPC (or MPC). The proposed RL-C-DMPC introduces RL modules to distributively obtain the compensations for the DG generation powers obtained from DMPC baseline control, effectively rectifying the power imbalances for all-electric diesel-solar vessels. Consequently, the frequency stability of the shipboard power systems can in the end be ensured.

The research on using RL to compensate for the control actions obtained by DMPC in distributed control frameworks has not been studied yet, so the proposed RL-C-DMPC is to explore this new area. Furthermore, we present a value-decomposition-network-based training and distributed testing mechanism while also proposing a method to select appropriate compensation rates tailored for shipboard power systems' energy management.

The effectiveness of RL-C-DMPC is tested through case studies based on real-life voyage data and historical solar generation power data. The results demonstrate that the proposed RL-C-DMPC substantially reduces the power imbalances, reaching around 90% reduction, when compared to the baseline control, i.e., DMPC. Furthermore, the proposed RL-C-DMPC, which has a distributed control framework, enhances the communication robustness compared to RL-C-MPC, which has a centralized control framework. This substantial reduction in power imbalances and enhancement in communication robustness indicate a promising approach to mitigate the influence of uncertain parameters in nominal prediction models in the energy management of shipboard power systems. The findings highlight the impacts of the proposed RL-C-DMPC strategy on control microgrids and multi-agent systems in distributed control frameworks since it does not need to formulate accurate physics-based nominal models for all uncertainty sources, making it a practical and efficient solution. In future works, we will include the compensations for charging/discharging powers in RL-C-DMPC.

References

[1] C. Park, B. Jeong, P. Zhou, H. Jang, S. Kim, H. Jeon, D. Nam, and A. Rashedi, "Live-life cycle assessment of the electric propulsion ship using solar PV," *Appl. Energy*, vol. 309, p. 118477, 2022.

[2] I. M. A. Nugraha, F. Luthfiani, G. Sotyaramadhani, A. Widagdo, and I. G. M. N. Desnanjaya, "Technical-economical assessment of solar pv systems on small-scale fishing vessels," *Int. J. Power Electron. Drive Syst*, vol. 13, no. 2, p. 1150, 2022.

[3] K. Wang, Y. Xue, H. Xu, L. Huang, R. Ma, P. Zhang, X. Jiang, Y. Yuan, R. R. Negenborn, and P. Sun, "Joint energy consumption optimization method for wing-diesel engine-powered hybrid ships towards a more energy-efficient shipping," *Energy*, vol. 245, p. 123155, 2022.

[4] F. Fan, V. Aditya, Y. Xu, B. Cheong, and A. K. Gupta, "Robustly coordinated operation of a ship microgrid with hybrid propulsion systems and hydrogen fuel cells," *Appl. Energy*, vol. 312, p. 118738, 2022.

[5] Y. Tao, J. Qiu, S. Lai, X. Sun, and J. Zhao, "Flexible voyage scheduling and coordinated energy management strategy of all-electric ships and seaport microgrid," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 3, pp. 3211–3222, 2022.

[6] M. Perčić, N. Vladimir, I. Jovanović, and M. Koričan, "Application of fuel cells with zero-carbon fuels in short-sea shipping," *Appl. Energy*, vol. 309, p. 118463, 2022.

[7] R. Zhang and H. Liang, "Application of solar energy in ship power field," in *2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*. IEEE, 2022, pp. 1588–1590.

[8] M. N. Nyanya, H. B. Vu, A. Schönborn, and A. I. Ölçer, "Wind and solar assisted ship propulsion optimisation and its application to a bulk carrier," *Sustain. Energy Technol. Assessments*, vol. 47, p. 101397, 2021.

[9] Z. Zhao, J. Guo, X. Luo, C. S. Lai, P. Yang, L. L. Lai, P. Li, J. M. Guerrero, and M. Shahidehpour, "Distributed robust model predictive control-based energy management strategy for islanded multi-microgrids considering uncertainty," *IEEE Transactions on Smart Grid*, vol. 13, no. 3, pp. 2107–2120, 2022.

[10] T. Morstyn and M. D. McCulloch, "Multiclass energy management for peer-to-peer energy trading driven by prosumer preferences," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 4005–4014, 2018.

[11] L. Olatomiwa, S. Mekhilef, M. S. Ismail, and M. Moghavvemi, "Energy management strategies in hybrid renewable energy systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 821–835, 2016.

[12] O. Jia-Richards and P. Lozano, "Analytical framework for staging of space propulsion systems," *Journal of Propulsion and Power*, vol. 36, no. 4, pp. 527–534, 2020.

[13] M. Bien, K. Ziaja, N. Blanken, Y. Cao, L. Schuchard, J. Göing, J. Friedrichs, F. di Mare, A. Mertens, B. Ponick *et al.*, "Modelling degradation mechanisms in hybrid-electric aircraft propulsion systems," in *25th International Symposium on Airbreathing Engines*, 2022.

[14] Y. Wang, M. Zechner, J. M. Mern, M. J. Kochenderfer, and J. K. Caers, "A sequential decision-making framework with uncertainty quantification for groundwater management," *Advances in Water Resources*, vol. 166, p. 104266, 2022.

[15] R. Gupta, F. Sossan, and M. Paolone, "Model-less robust voltage control in active distribution networks using sensitivity coefficients estimated from measurements," *Electric Power Systems Research*, vol. 212, p. 108547, 2022.

[16] A. C. Caputo, A. Federici, P. M. Pelagagge, and P. Salini, "Offshore wind power system economic evaluation framework under aleatory and epistemic uncertainty," *Applied Energy*, vol. 350, p. 121585, 2023.

[17] J. Ding, K. Xie, B. Hu, C. Shao, T. Niu, C. Li, and C. Pan, "Mixed aleatory-epistemic uncertainty modeling of wind power forecast errors in operation reliability evaluation of power systems," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 5, pp. 1174–1183, 2022.

[18] Y. Huang, Y. Wang, and N. Liu, "A two-stage energy management for heat-electricity integrated energy system considering dynamic pricing of stackelberg game and operation strategy optimization," *Energy*, vol. 244, p. 122576, 2022.

[19] X. Zhu, Y. Sun, J. Yang, Z. Dou, G. Li, C. Xu, and Y. Wen, "Day-ahead energy pricing and management method for regional integrated energy systems considering multi-energy demand responses," *Energy*, vol. 251, p. 123914, 2022.

[20] H. Pan, X. Chen, T. Jin, Y. Bai, Z. Chen, J. Wen, and Q. Wu, "Real-time power market clearing model with improved network constraints considering PTDF correction and fast-calculated dynamic line rating," *IEEE Transactions on Industry Applications*, vol. 59, no. 2, pp. 2130–2139, 2022.

[21] S.-H. Hong and H.-S. Lee, "Robust energy management system with safe reinforcement learning using short-horizon forecasts," *IEEE Trans. Smart Grid*, vol. 14, no. 3, pp. 2485–2488, 2023.

[22] R. Lu, Z. Jiang, H. Wu, Y. Ding, D. Wang, and H.-T. Zhang, "Reward shaping-based actor–critic deep reinforcement learning for residential energy management," *IEEE Trans. Industr. Inform.*, vol. 19, no. 3, pp. 2662–2673, 2023.

[23] C. Huang, H. Zhang, L. Wang, X. Luo, and Y. Song, "Mixed deep reinforcement learning considering discrete-continuous hybrid action space for smart home energy management," *J. Mod. Power Syst. Clean Energy*, vol. 10, no. 3, pp. 743–754, 2022.

[24] H. H. Goh, Y. Huang, C. S. Lim, D. Zhang, H. Liu, W. Dai, T. A. Kurniawan, and S. Rahman, "An assessment of multistage reward function design for deep reinforcement learning-based microgrid energy management," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4300–4311, 2022.

[25] Y. Hao, Q. Lu, X. Wang, and B. Jiang, "Adaptive model-based reinforcement learning for fast charging optimization of Lithium-Ion batteries," *IEEE Trans. Industr. Inform.*, pp. 1–10, 2023, doi: 10.1109/TII.2023.3257299.

[26] Z. Yi, Y. Xu, and C. Wu, "Model-free economic dispatch for virtual power plants: An adversarial safe reinforcement learning approach," *IEEE Trans. Power Syst.*, pp. 1–15, 2023, doi: 10.1109/TPWRS.2023.3289334.

[27] Y. Du and F. Li, "Intelligent multi-microgrid energy management based on deep neural network and model-free reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1066–1076, 2020.

[28] Y. Li, R. Wang, and Z. Yang, "Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 159–169, 2022.

[29] K. Ojand and H. Dagdougui, "Q-learning-based model predictive control for energy management in residential aggregator," *IEEE Trans. Autom. Sci. Eng*, vol. 19, no. 1, pp. 70–81, 2022.

[30] J. Wang, J. Wu, and X. Kong, "Multi-agent simulation for strategic bidding in electricity markets using reinforcement learning," *CSEE J. Power Energy Syst.*, vol. 9, no. 3, pp. 1051–1065, 2023.

[31] Y. Zhang, H. Wen, Q. Wu, and Q. Ai, "Optimal adaptive prediction intervals for electricity load forecasting in distribution systems via reinforcement learning," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 3259–3270, 2023.

[32] Q. Gao, Y. Liu, J. Zhao, J. Liu, and C. Y. Chung, "Hybrid deep learning for dynamic total transfer capability control," *IEEE Trans. Power Syst.*, vol. 36, no. 3, pp. 2733–2736, 2021.

[33] Q. Zhang, W. Pan, and V. Reppa, "Model-reference reinforcement learning for collision-free tracking control of autonomous surface vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8770–8781, 2021.

[34] Y. Lu, C. Wu, W. Yao, G. Sun, J. Liu, and L. Wu, "Deep reinforcement learning control of fully-constrained cable-driven parallel robots," *IEEE Trans. Ind. Electron.*, 2022.

[35] W. Remmerswaal, D. Sun, A. Jamshidnejad, and B. De Schutter, "Combined MPC and reinforcement learning for traffic signal control in urban traffic networks," in *2022 26th International Conference on System Theory, Control and Computing (ICSTCC)*. IEEE, 2022, pp. 432–439.

[36] J. S. Gomez, D. Saez, J. W. Simpson-Porco, and R. Cárdenas, "Distributed predictive control for frequency and voltage regulation in

microgrids," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1319–1329, 2019.

[37] P. Kou, D. Liang, and L. Gao, "Distributed coordination of multiple PMSGs in an islanded DC microgrid for load sharing," *IEEE Transactions on Energy Conversion*, vol. 32, no. 2, pp. 471–485, 2017.

[38] A. Parisio, C. Wiezorek, T. Kyntäjä, J. Elo, K. Strunz, and K. H. Johansson, "Cooperative MPC-based energy management for networked microgrids," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 3066–3074, 2017.

[39] H. Xiao, X. Pu, W. Pei, L. Ma, and T. Ma, "A novel energy management method for networked multi-energy microgrids based on improved DQN," *IEEE Trans. Smart Grid*, pp. 1–1, 2023, doi: 10.1109/TSG.2023.3261979.

[40] H. Xiao, L. Fu, C. Shang, X. Bao, X. Xu, and W. Guo, "Ship energy scheduling with DQN-CE algorithm combining bi-directional LSTM and attention mechanism," *Appl. Energy*, vol. 347, p. 121378, 2023.

[41] H. Zeng, B. Shao, H. Dai, N. Tian, and W. Zhao, "Incentive-based demand response strategies for natural gas considering carbon emissions and load volatility," *Appl. Energy*, vol. 348, p. 121541, 2023.

[42] P. Sunehag, G. Lever, A. Gruslys, W. M. Czarnecki, V. Zambaldi, M. Jaderberg, M. Lanctot, N. Sonnerat, J. Z. Leibo, K. Tuyls *et al.*, "Value-decomposition networks for cooperative multi-agent learning," *arXiv preprint arXiv:1706.05296*, 2017.

[43] Z. Guo, Y. Wu, L. Wang, and J. Zhang, "Coordination for connected and automated vehicles at non-signalized intersections: A value decomposition-based multiagent deep reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3025–3034, 2023.

[44] Z. Qiu, C. He, and X. Zhang, "Multi-agent cooperative structural vibration control of three coupled flexible beams based on value decomposition network," *Eng. Appl. Artif. Intell.*, vol. 114, p. 105002, 2022.

[45] T. Cao and X. Zhang, "Nonlinear decoration control based on perturbation of ship longitudinal motion model," *Applied Ocean Research*, vol. 130, p. 103412, 2023.

[46] J. Hou, J. Sun, and H. Hofmann, "Adaptive model predictive control with propulsion load estimation and prediction for all-electric ship energy management," *Energy*, vol. 150, pp. 877–889, 2018.

[47] P. Xie, S. Tan, N. Bazmohammadi, J. M. Guerrero, J. C. Vasquez, J. M. Alcala, and J. E. M. Carreño, "A distributed real-time power management scheme for shipboard zonal multi-microgrid system," *Appl. Energy*, vol. 317, p. 119072, 2022.

[48] Ø. N. Smogeli, *Control of marine propellers: from normal to extreme conditions*. Fakultet for ingeniørvitenskap og teknologi, 2006.

[49] S. Nasiri, S. Peyghami, M. Parniani, and F. Blaabjerg, "Modeling in-and-out-of-water impact on all-electric ship power system considering propeller submergence in waves," in *2021 IEEE Transportation Electrification Conference & Expo (ITEC)*. IEEE, 2021, pp. 533–538.

[50] M. Mohamed, F. E. Mahmood, M. A. Abd, A. Chandra, and B. Singh, "Dynamic forecasting of solar energy microgrid systems using feature engineering," *IEEE Trans. Ind. Appl.*, vol. 58, no. 6, pp. 7857–7869, 2022.

[51] D. van der Meer, G. R. Chandra Mouli, G. Morales-España Mouli, L. R. Elizondo, and P. Bauer, "Energy management system with PV power forecast to optimally charge EVs at the workplace," *IEEE Trans. Industr. Inform.*, vol. 14, no. 1, pp. 311–320, 2018.

[52] P. Gangwar, A. Mallick, S. Chakrabarti, and S. N. Singh, "Short-term forecasting-based network reconfiguration for unbalanced distribution systems with distributed generators," *IEEE Trans. Industr. Inform.*, vol. 16, no. 7, pp. 4378–4389, 2019.

[53] A. Jamshidnejad, D. Sun, A. Ferrara, and B. De Schutter, "A novel bi-level temporally-distributed MPC approach: An application to green urban mobility," *Available at SSRN 4370158*.

[54] Y. Wang, L. Wu, and S. Wang, "A fully-decentralized consensus-based admm approach for dc-opf with demand response," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2637–2647, 2016.

[55] A. Korompili, P. Pandis, and A. Monti, "Distributed OPF algorithm for system-level control of active multi-terminal DC distribution grids," *IEEE Access*, vol. 8, pp. 136 638–136 654, 2020.

[56] M. A. Mohamed, H. Chabok, E. M. Awwad, A. M. El-Sherbeeny, M. A. Elmeligy, and Z. M. Ali, "Stochastic and distributed scheduling of shipboard power systems using MθFOA-ADMM," *Energy*, vol. 206, p. 118041, 2020.

[57] Z. Zhang, W. Tian, and Z. Liao, "Towards coordinated and robust real-time control: A decentralized approach for combined sewer overflow and urban flooding reduction based on multi-agent reinforcement learning," *Water Research*, vol. 229, p. 119498, 2023.

[58] J. Zhao, X. Hu, M. Yang, W. Zhou, J. Zhu, and H. Li, "Ctds: Centralized teacher with decentralized student for multi-agent reinforcement learning," *IEEE Transactions on Games*, 2022.

[59] V. Vittal, J. McCalley, P. M. Anderson, and A. Fouad, *Power System Control and Stability*. John Wiley & Sons, 2019.

[60] T. S. Mummadi and V. R., "Optimal design and power management in shipboard system," *CVR journal of science and technology*, vol. 17, pp. 83–89, 04 2020.

[61] S. Wen, H. Lan, D. C. Yu, Q. Fu, Y. Y. Hong, L. Yu, and R. Yang, "Optimal sizing of hybrid energy storage sub-systems in PV/Diesel ship power system using frequency analysis," *Energy*, vol. 140, no. pt.1, pp. 198–208, 2017.

[62] J. Xu, H. Sun, and C. J. Dent, "ADMM-based distributed OPF problem meets stochastic communication delay," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5046–5056, 2019.

[63] J. Guo, G. Hug, and O. Tonguz, "Impact of communication delay on asynchronous distributed optimal power flow using ADMM," in *2017 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2017, pp. 177–182.

[64] C. H. Ho, H. C. Wu, S. C. Chan, and Y. Hou, "A robust statistical approach to distributed power system state estimation with bad data," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 517–527, 2020.

[65] M. H. Nazari, L. Y. Wang, S. Grijalva, and M. Egerstedt, "Communication-failure-resilient distributed frequency control in smart grids: Part I: Architecture and distributed algorithms," *IEEE Transactions on Power Systems*, vol. 35, no. 2, pp. 1317–1326, 2020.