12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Greenhouse Gas Emissions Reduction of a Hybrid-Powered Ferry using Deep Reinforcement Learning for Power Load Distribution ⋆

A. Abdalla * **B. Gopaluni** ** **P. Kirchen** ***

* *Chemical & Biological Engineering Department, University of British Columbia, Vancouver, BC V6T 1Z3 Canada, (e-mail: ahmedoa@mail.ubc.ca).*
** *Chemical & Biological Engineering Department, University of British Columbia, Vancouver, BC V6T 1Z3 Canada, (e-mail: bhushan.gopaluni@ubc.ca)*
*** *Mechanical Engineering Department, University of British Columbia, Vancouver, BC V6T 1Z4 Canada, (e-mail: pkirchen@mech.ubc.ca)*

Abstract: This article explores the use of the twin delayed deep deterministic policy gradient (TD3), a deep reinforcement learning algorithm, to reduce the cumulative greenhouse gas (GHG) emissions from the sailing trips of a hybrid-powered roll-on roll-off liquefied natural gas ferry. The objective of the algorithm is to optimally control the power load distribution between the ferry's engines and battery to achieve a reduction in GHG emissions. Results from this study show that the TD3 agent achieved an average reduction in cumulative GHG emissions by 5% against actual operations for the sailing trips that were analyzed. The performance of the TD3 agent was compared to a rule-based energy management strategy (EMS) in which the ferry's battery was used to operate the ferry completely at low load operations and provide surplus power when the power demand was greater than the engine rating. The rule-based EMS failed to achieve GHG emissions reductions against actual operations thereby indicating the efficacy of the TD3 agent in achieving emissions reductions.

Keywords: Artificial intelligence in transportation, Deep reinforcement learning, GHG reduction, LNG ferry operation optimization, Power load distribution control.

1. INTRODUCTION

The International Maritime Organization (IMO) has set ambitious targets in the 2023 IMO strategy on reduction of greenhouse gas (GHG) emissions from ships. The aim is for international shipping to reach net-zero GHG emissions by 2050. The strategy also outlines checkpoints to reach the net-zero emissions, which include a reduction in annual GHG emissions by 20-30% by 2030, and a reduction of 70-80% by 2040 (IMO, 2023). Both the estimates are compared to annual emissions in 2008. Emissions from global shipping is estimated to increase by 23% by 2035 compared to 2015 if no additional policy measures are established (IMO, 2018). Based on current technologies which include options such as alternative fuels, hybridization, and optimization of operations, there is a potential to reduce emissions from maritime transportation by 75-85% (Bouman et al., 2017). Liquefied natural gas (LNG) is a promising fuel for use in the efforts to decarbonize the maritime industry. However, the carbon dioxide reduction achieved by using LNG as a fuel is offset by high methane emissions at low engine operational loads, known as methane slips, due to factors such as incomplete combustion and low flame speeds (Bouman et al., 2017; Sommer et al., 2019; Rochussen et al., 2023; Balcombe et al., 2022). The design of an energy management strategy (EMS) to efficiently use the LNG engines is therefore mandatory to realize the true GHG emissions reduction potential of using LNG as a fuel. The EMS is used to optimally control the distribution of power between the components of a power system given a prespecified objective.

There are a range of EMSs that are employed for the optimization of hybrid vessel powertrains including rule-based, optimization-based, and learning-based approaches (Cha et al., 2023). Studies implemented in literature on the EMS of hybrid-powered LNG vessels focused on the use of rule-based and optimization-based approaches to reduce cumulative GHG emissions from sailing trips (Feng et al., 2023; Fan et al., 2023; Roslan et al., 2023; Stamatakis and Ioannides, 2021). Most of the studies, however, only considered carbon dioxide in the GHG emissions reduction objective and not methane emissions.

Learning-based approaches in the form of reinforcement learning (RL) algorithms have been used to minimize the operational costs of non-LNG hybrid-powered ves-

⋆ Funded by the Climate Action and Awareness Fund (CAAF).

sels through power load distribution control. Successful applications include the use of a Double Q agent (Wu et al., 2020), Deep Q Network (Jung and Chang, 2023; Shang et al., 2022), Double Deep Q Network (Wu et al., 2021b) and the twin delayed deep deterministic policy gradient (TD3) (Wu et al., 2021a). The RL algorithms were trained using historical power demand profiles of hybrid-powered vessels and the cost minimization objective was set through the reward functions used in the algorithms. The cost minimization performance of the RL algorithms was however only compared to the performance of other baseline algorithms that are considered to provide optimal performance.

This study aims to determine the feasibility of using deep RL in optimizing the power load distribution of a hybrid-powered LNG ferry. The optimization goal is a reduction in the ferry's cumulative GHG emissions at the end of each sailing trip while incorporating both carbon dioxide and methane in the emissions reduction objective. TD3 is the deep RL algorithm used in this study. TD3 was chosen as it works with continuous action spaces and was found to outperform other state-of-the-art RL algorithms when tested using multiple model environments (Fujimoto et al., 2018).

2. DEEP REINFORCEMENT LEARNING AND ACTOR-CRITIC METHODS

Deep RL is a combination of deep learning and RL in which a deep network is used to estimate any of the RL components; namely, the value function, policy, state transition function or the reward function (Li, 2018). Actor-critic methods are those that learn approximations to both the policy and value functions. In actor-critic methods, the actor refers to the learned policy that is used by the RL agent to select actions, whereas the critic refers to the learned value function that evaluates the actions taken by the agent (Sutton and Barto, 2018).

2.1 TD3 Algorithm

TD3 is an actor-critic deep RL algorithm developed by Fujimoto et al. (2018) that is designed to work with continuous action spaces. TD3 was developed as an improvement to the Q-learning based algorithms that have been found to suffer from state value overestimation bias and learning suboptimal policies. The improvement in performance by TD3 is achieved by:

- Target policy smoothing: A noise is added to the target policy to mitigate the overestimation that occurs when calculating the target Q-value.
- Target policy delay: The update of the target policy is delayed with respect to the update of the target Q-network in order to reduce the for-every-update error and improve performance.
- Clipped twin critic networks: Two critic networks are used instead of one. The lower of the two Q-values calculated by the critic networks is used for network updates to reduce Q-value overestimation.

2.2 Objective of the TD3 Agent

The objective of the TD3 agent in the current analysis is to control the distribution of power between the LNG engines and the battery system of a hybrid-powered LNG ferry such that minimum cumulative GHG emissions for sailing trips are attained. The ferry under study is a roll-on roll-off ferry that operates in the Canadian Salish Sea.

GHG emissions were chosen as the objective of optimization instead of fuel consumption due to the increase in the brake-specific GHG emission factors of the ferry's engine at low loads. The increase in the brake-specific GHG emission factors is typical of four-stroke low-pressure dual-fuel engines at low loads. A lower fuel consumption is not directly correlated to lower GHG emissions in carbon dioxide equivalent (CO_{2e}).

The performance of the TD3 agent is compared to that of a rule-based approach which uses the following strategy:

- The ferry is operated completely using battery power when the power demand is ≤800 kW (≤16.7% engine load).
- The battery is used to supply extra required power when the power demand >4770 kW (>100% engine load).

3. MODEL ENVIRONMENT

The characteristics of the ferry under study are summarized in Table 1. These characteristics were used to define the model environment used to train the TD3 agent.

Table 1. Characteristics for the ferry under study.

Characteristic	Value
Vessel type	Roll-on roll-off cargo ferry
Build year	2021
Installed engine power	2x 4770 kW
Fueling	Low-pressure dual-fuel
Battery capacity	2034 kWh
Maximum battery charge rate	2000 kW
Maximum battery discharge rate	800 kW

Operating the ferry using one engine instead of two leads to lower attained cumulative GHG emissions (Rochussen et al., 2023; Sommer et al., 2019). The RL model environment has therefore been setup to work with one engine.

The cumulative GHG emissions from the sailing trips is the sum of GHG emissions from two operations: GHG emissions released during the sailing period and GHG emissions released when recharging the battery back to a nominal state of charge (SOC) of 85%. The battery is recharged at the end of the trips using the ferry's LNG engine at a power load of 2000 kW. The GHG emission rates were calculated using steady-state emission factors developed by Rochussen et al. (2023) that estimate the instantaneous GHG emission rates as a function of the engine load. The emission factors are based on 1-minute averages of emission rate measurements at fixed engine loads.

3.1 Data Preparation

Data collected from the ferry includes historical engine and battery power profiles, travel time, and battery SOC. The required data was collected using MODBUS signals at a 1 Hz frequency over 14 sailing trips. Before model training, each signal was converted into 1-minute averages to reduce noise and match the time resolution used in the estimation of the ferry's GHG emission factors.

The total power demand of the ferry was calculated for each minute by adding the battery power to the engine power (1).

$$P_{tot,t} = P_{bat,t} + P_{eng,t} \tag{1}$$

where $P_{tot,t}$ is the total power demand, $P_{bat,t}$ is the battery power, and $P_{eng,t}$ is the engine power at time t.

The total power demand and engine power were normalized by their respective maximum values. The power demand profiles were used to train and test the TD3 agent. A total of 14 power demand profiles were available. Ten profiles were used to train the TD3 agent while four profiles were used to test the agent's performance on unseen power demand profiles.

In the context of this article, *actual* conditions refer to values obtained from the minute-averaged raw data of sailing trips (i.e., not generated by the TD3 agent or rule-based strategy).

3.2 State Space

The state space (2) of the RL model environment consists of the fraction of trip completion (current time/max trip time), normalized power demand, engine load (normalized engine power), and battery SOC fraction. The state space is a 4-dimensial continuous array.

$$s_t = [T_{frac,t}, P_{dem,t}, L_t, C_t]^T \tag{2}$$

where s_t is the state space array at time t, $T_{frac,t}$ is the fractional trip completion, $P_{dem,t}$ is the normalized power demand, L_t is the engine load, and C_t is the battery SOC fraction.

3.3 Action Space

The action space (3) of the RL model environment consists of the change in engine load. The engine load change for each time step is a continuous 1-dimensional array between -0.5 and 0.5.

$$A = [\Delta L] \tag{3}$$

where ΔL is the change in engine load per minute.

The engine load at the next state is calculated by adding the change in engine load chosen by the TD3 agent to the engine load in the current state. If an action is chosen by the agent such that the engine load becomes negative or greater than 1, the model's environment clips the output to keep it within the specified limits (4).

$$L_{t+1} = \begin{cases} 0 & \text{if } L_t + \Delta L_t < 0, \\ 1 & \text{if } L_t + \Delta L_t > 1, \\ L_t + \Delta L_t & \text{else.} \end{cases} \tag{4}$$

where L_{t+1} is the engine load in state s_{t+1}, L_t is the engine load in state s_t, and ΔL_t is the change in engine load chosen by the agent at state s_t.

The battery SOC is calculated using (5).

$$C_{t+1} = C_t + \frac{1}{Q_{\text{cap}}} \int_t^{t+1} \dot{Q}\, dt \tag{5}$$

where Q_{cap} is the total battery capacity corrected for its state of health and $\dot{Q}$ is the battery charge rate corrected for efficiency.

3.4 Reward Function

The reward function was designed to incentivize the TD3 agent to learn a policy that minimizes cumulative GHG emissions. Equations (6)-(8) were used to calculate the GHG reward between each time step. The reward approaches a value of 0 for state transitions that lead to high GHG emissions and approaches a value of 1 for state transitions that lead to low GHG emissions.

The GHG reward is defined as:

$$GHG_{inst} = -22603L^4 + 74208L^3 - 77912L^2 + 29979.3L \tag{6}$$

$$\Delta GHG_{t+1} = \int_t^{t+1} GHG_{inst}\ dt \tag{7}$$

$$r_{ghg,t+1} = exp(\frac{-\Delta GHG_{t+1}}{40}) \tag{8}$$

where GHG_{inst} is the instantaneous GHG emission rate in $kg\ CO_{2e}$, L is the engine load, ΔGHG_{t+1} is the amount of GHG released within one timestep, and $r_{ghg,t+1}$ is the GHG reward. The factor of 40 in (8) was used to scale the GHG reward between 0 and 1.

Negative engine loads that are encountered during the training of the TD3 agent were transformed to a value of 1 within the calculation of the GHG reward. The transformation of the negative load to a larger value leads to a diminished GHG reward and disincentivizes the agent from taking actions that lead to negative engine loads. The transformation maintains the GHG reward between 0 and 1.

Equation (9) was used to calculate the terminal GHG reward penalty that considers GHG emissions released when recharging the battery back to an SOC of 85%.

$$r_{ghg,term} = \frac{-\Delta GHG_{ch}}{20} \tag{9}$$

where $r_{ghg,term}$ is the terminal GHG penalty and ΔGHG_{ch} is the amount of GHG released when recharging the battery back to an SOC of 85% at the end of a sailing trip. The

factor of 20 was a tuned hyperparameter used to adjust the scale of the reward.

The agent was penalized with a reward of -1 for state transitions that led to unviable states such as:

- Engine loads beyond the thresholds of 0 to 1 (10).
- Battery SOC fraction that is greater than 0.85 or less than 0.30 (11). This was selected as per the ferry operator's battery usage philosophy.
- Battery power beyond the thresholds of -2000 to 800 kW (12). This was selected as per the battery power limits. Negative power values indicate battery charging while positive power values indicate battery discharging.

$$r_{L,t+1} = \begin{cases} -1 & \text{if } L_t + \Delta L_t \notin [0,1], \\ 0 & \text{else.} \end{cases} \tag{10}$$

$$r_{soc,t+1} = \begin{cases} -1 & \text{if } C_{t+1} \notin [0.30, 0.85], \\ 0 & \text{else.} \end{cases} \tag{11}$$

$$r_{bat,t+1} = \begin{cases} -1 & \text{if } P_{bat,t+1} \notin [-2000, 800] kW, \\ 0 & \text{else.} \end{cases} \tag{12}$$

where $r_{L,t+1}$ is the engine load penalty, $r_{soc,t+1}$ is the battery SOC penalty, and $r_{bat,t+1}$ is the battery power penalty.

The total reward function is therefore defined as:

$$R_{t+1} = \begin{cases} r_{ghg,t+1} + r_{L,t+1} + r_{soc,t+1} + r_{bat,t+1} \\ \quad \text{if } T_{\text{frac},t} < 1, \\ r_{ghg,t+1} + r_{L,t+1} + r_{soc,t+1} + r_{bat,t+1} + r_{ghg,\text{term}} \\ \quad \text{if } T_{\text{frac},t} = 1. \end{cases} \tag{13}$$

where R_{t+1} is the total reward at the end of each timestep.

4. MODEL TRAINING

This section describes the hardware, libraries and hyperparameter configurations that were used to train the TD3 agent.

4.1 Hardware Configuration

The TD3 agent was trained using a Lenovo ideapad gaming III laptop utilizing an AMD Ryzen 5000 series CPU with a base processor speed of 3.30 GHz and 16 gb of RAM.

4.2 Model Setup

The RL environment was set up using OpenAI's *gym* library. The TD3 agent was set up using the *stable_baselines*3 library. The hyperparameters of the TD3 agent were tuned using the *Optuna* library. The hyperparameter configuration used to train the TD3 agent is summarized in Table 2.

Gaussian action noise with linear decay was added to the input of the TD3 agent to encourage exploration. The hyperparameter configuration of the noise is summarized in Table 3.

Table 2. TD3 agent hyperparameter settings.

Hyperparameter	Value
Learning Rate, α	0.00125
Discount Factor, γ	1
Soft Update Coefficient, τ	0.05
Batch Size	256
Buffer Size	1×10^6
Actor Network Size	[500, 400]
Critic Network Size	[500, 400]
Training Frequency	16 steps

Table 3. Gaussian noise with linear decay hyperparameter settings.

Hyperparameter	Value
Mean, μ	0.00
Initial Standard Deviation, σ_0	0.66
Maximum Decay Steps	50,000
Final Standard Deviation, σ_f	0.06

4.3 Training Configuration

One training episode consists of the measured power demand profile of one sailing trip. At each time step, the following sequence takes place:

- The TD3 agent decides the best change in engine load using information only from the current state.
- The engine load in the next state is calculated using (4).
- The required battery power in the next state is calculated using (1) given the power demand and calculated engine load for the next state.
- The battery SOC in the next state is calculated using (5).
- The time is incremented by 1 minute.

The episode is terminated when $T_{\text{frac},t}$ in the state space is equal to 1. During each training episode iteration, a power demand profile from the ten that are available for training was randomly selected to train the TD3 agent. The agent loops through the ten power demand profiles until a policy is learned which leads to the convergence of the total reward signal received by the agent.

The initial state of each training episode was set as follows:

- The fractional trip completion and power demand are the initial values from the randomly selected power demand profile (i.e., when $T_{\text{frac},t} = 0$).
- The initial engine load is 0.
- The initial battery SOC is a random number between 0.75 and 0.85. This was configured to increase the variability of starting states.

4.4 Training Performance

The TD3 agent was trained for 100,000 timesteps ($\sim$470 episodes). The training time was $\sim$13 minutes. The average total reward per episode was logged at an interval of 10 episodes. A total of 5 separate training iterations were run with different random seeds to ensure repeatability of obtained results.

As seen in Fig. 1, there was sharp learning between episodes 0 and 300. Convergence and the maximum total reward per episode was achieved after $\sim$340 episodes. The

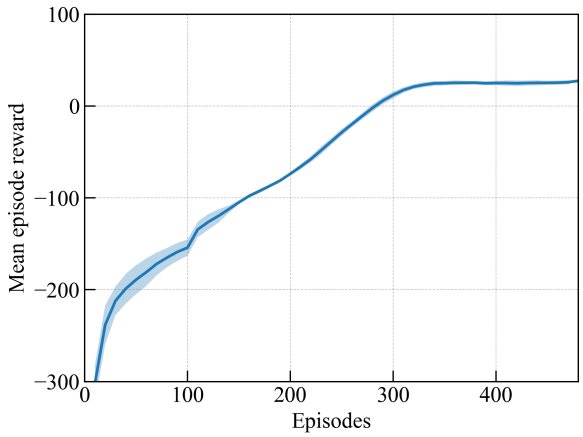

Fig. 1. Learning curve of the TD3 agent during training.

shaded region represents one standard deviation over the rewards from the five different training iterations.

5. MODEL PERFORMANCE COMPARISON

Engine load profiles for each power demand profile were obtained using outputs from the TD3 agent. The engine load profiles were converted to engine power (kW) by multiplying each value by the engine's power rating. Battery power profiles (kW) were calculated using (1). Outputs by the TD3 agent that led to engine load or battery power values beyond the thresholds were clipped to remain within the thresholds. For the purpose of evaluation, the initial battery SOC for each trip provided to the TD3 agent was the actual initial battery SOC of each trip.

Fig. 2 shows a comparison between the TD3 agent's output and the actual power load distribution for two sample sailing trips along with their respective battery SOC curves. Fig. 2 (a) represents a sailing trip that was part of the training dataset whereas Fig. 2 (b) represents a sailing trip that was part of the test dataset. Fig. 3 illustrates the power and battery SOC profiles generated following the rule-based strategy and compared against the actual profiles of a sample sailing trip.

5.1 Estimation of Cumulative GHG Emissions

The cumulative GHG emissions from the actual trips, TD3 generated profiles and the profiles generated using the rule-based approach were calculated using (6). The cumulative GHG emissions at the end of each trip is composed of the integrated instantaneous GHG emission rate over each trip's time duration plus the amount of GHG emissions released when recharging the battery to an SOC of 85% using the engine at a power load of 2000 kW.

Table 4 summarizes the performance of both the TD3 agent and rule-based strategy in terms of average cumulative GHG emissions over the sailing trips and the percentage emissions reductions achieved.

5.2 Results Summary

The TD3 agent was successful in providing engine and battery SOC profiles that are within the operational constraints of the ferry as seen in Fig. 2. The actual and TD3-generated battery power profiles display significant differences primarily at the start and the end of the sailing trips. The observed differences in the battery power profiles are attributed to the TD3 agent's attempts in minimizing cumulative GHG emissions by optimizing the battery's charge and discharge cycles. The battery SOC profiles generated following the rule-based strategy, as seen in the example in Fig. 3, were maintained within the battery SOC thresholds. However, significant charging was required at the end of the trips as no charging occurred during the sailing periods.

The average reduction in cumulative GHG emissions achieved by the TD3 agent was 4% when used on the training dataset and 6% when used on the test dataset. The average emissions reduction achieved by the TD3 agent for the entire dataset was 5%. The results indicate that the TD3 agent learned a good policy to choose engine loads that minimize GHG emissions and that the agent can operate on sailing trips it has not been trained on. The average emissions reduction achieved on the test dataset was slightly higher as compared to the training dataset as one of the trips in the test dataset had more potential for GHG reduction compared to the rest of the trips.

The rule-based EMS led to almost no reduction in average GHG emissions when compared to actual conditions. The performance of the TD3 agent was therefore better as compared to the rule-based strategy. The failure of the rule-based strategy indicates that the rules of switching to complete battery usage at low engine loads and using the battery when the power demand is higher than the engine rating are not adequate to minimize cumulative GHG emissions from sailing trips.

6. CONSLUSION

The TD3 agent used in this study was successful in reducing the cumulative GHG emissions for the trips that were evaluated. The TD3 agent led to a reduction in GHG emissions by 5% on average over all the sailing trips evaluated. The rule-based strategy, on the other hand, failed to reduce the GHG emissions of the sailing trips. The GHG reductions were achieved by the TD3 agent while complying with the ferry's operational constraints.

To the best of the authors' knowledge, the application of deep RL to minimize cumulative GHG emissions from hybrid-powered LNG ferries has not been explored in existing literature. This research aims to fill this gap by studying the feasibility of using deep RL techniques in reducing GHG emissions from this specific maritime context.

The future work in the course of this research is to:

- Improve the performance of the TD3 agent to achieve more cumulative GHG reductions for sailing trips by incorporating more training data.
- Compare the performance of the TD3 agent to other state-of-the-art actor-critic RL algorithms.
- Assess the potential of using deep RL in real-time power load distribution control to reduce cumulative GHG emissions from sailing trips.

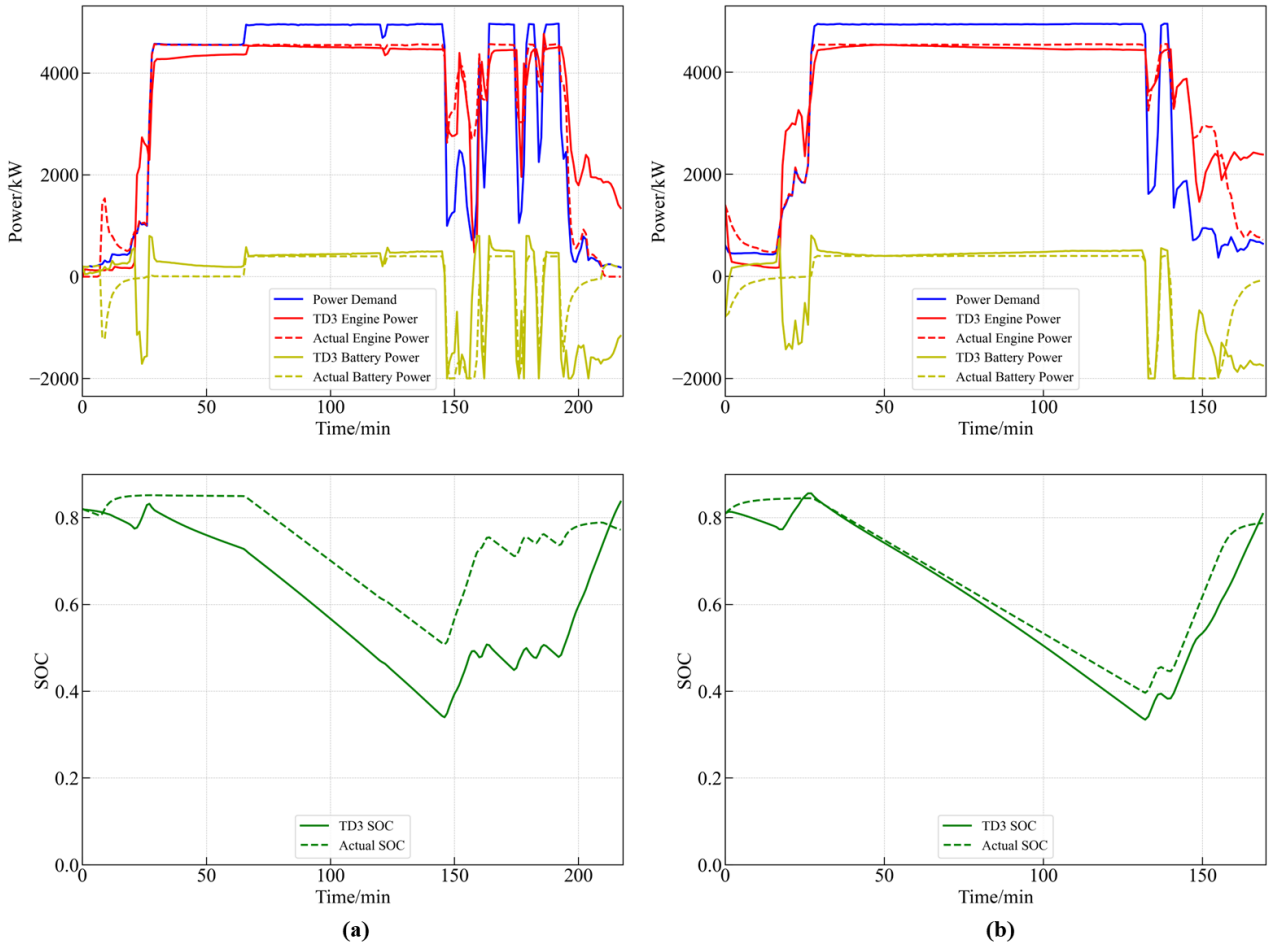

Fig. 2. Plots of the actual vs TD3 output power and battery SOC profiles for (a) a sailing trip from the training dataset and (b) a sailing trip from the test dataset.

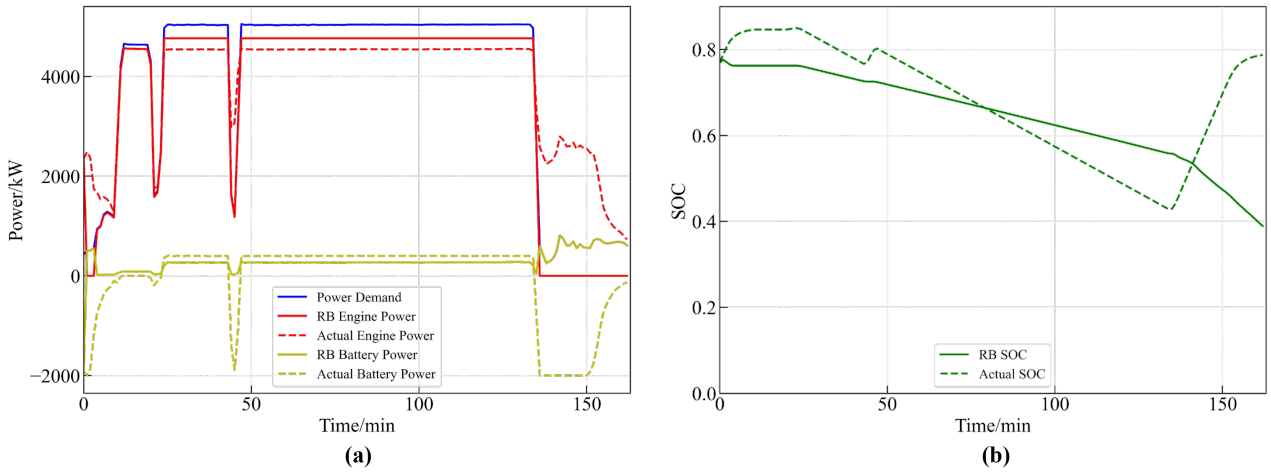

Fig. 3. Plots of the actual vs rule-based (RB) output power and battery SOC profiles for a sample sailing trip.

Table 4. Average cumulative GHG emissions and GHG emissions reduction summary comparing the performance of an EMS using a TD3 agent and an EMS using a rule-based approach.

Dataset	Actual GHG Emissions ($kg\ CO_{2e}$)	TD3 Optimized GHG Emissions ($kg\ CO_{2e}$)	Rule-Based GHG Emissions ($kg\ CO_{2e}$)	TD3 GHG Reduction (%)	Rule-Based GHG Reduction (%)
Training	11081	10612	11075	4	0
Test	10374	9708	10328	6	0
Overall	10863	10334	10845	5	0

REFERENCES

Balcombe, P., Heggo, D.A., and Harrison, M. (2022). Total methane and co2 emissions from liquefied natural gas carrier ships: The first primary measurements. *Environmental Science & Technology*, 56(13), 9632–9640. doi:10.1021/acs.est.2c01383. URL https://doi.org/10.1021/acs.est.2c01383. PMID: 35699220.

Bouman, E.A., Lindstad, E., Rialland, A.I., and Strømman, A.H. (2017). State-of-the-art technologies, measures, and potential for reducing ghg emissions from shipping–a review. *Transportation Research Part D: Transport and Environment*, 52, 408–421.

Cha, M., Enshaei, H., Nguyen, H., and Jayasinghe, S. (2023). Towards a future electric ferry using optimisation-based power management strategy in fuel cell and battery vehicle application — a review. *Renewable and Sustainable Energy Reviews*, 183, 113470. doi:https://doi.org/10.1016/j.rser.2023.113470.

Fan, A., Li, Y., Fang, S., Li, Y., and Qiu, H. (2023). Energy management strategies and comprehensive evaluation of parallel hybrid ship based on improved fuzzy logic control. *IEEE Transactions on Transportation Electrification*, 1–1. doi:10.1109/TTE.2023.3332772.

Feng, Y., Zhu, H., and Dong, Z. (2023). Simultaneous and global optimizations of lng-fueled hybrid electric ship for substantial fuel cost, co2, and methane emission reduction. *IEEE Transactions on Transportation Electrification*, 9(2), 2282–2295. doi:10.1109/TTE.2022.3208880.

Fujimoto, S., Hoof, H., and Meger, D. (2018). Addressing function approximation error in actor-critic methods. In *International conference on machine learning*, 1587–1596. PMLR.

IMO (2018). Decarbonising maritime transport: Pathways to zero-carbon shipping by 2035.

IMO (2023). 2023 imo strategy on reduction of ghg emissions from ships.

Jung, W. and Chang, D. (2023). Deep reinforcement learning-based energy management for liquid hydrogen-fueled hybrid electric ship propulsion system. *Journal of Marine Science and Engineering*, 11(10), 2007.

Li, Y. (2018). Deep reinforcement learning: An overview.

Rochussen, J., Jaeger, N.S., Penner, H., Khan, A., and Kirchen, P. (2023). Development and demonstration of strategies for ghg and methane slip reduction from dual-fuel natural gas coastal vessels. *Fuel*, 349, 128433.

Roslan, S.B., Tay, Z.Y., Konovessis, D., Ang, J.H., and Menon, N.V. (2023). Rule-based control studies of lng–battery hybrid tugboat. *Journal of Marine Science and Engineering*, 11(7). doi:10.3390/jmse11071307. URL https://www.mdpi.com/2077-1312/11/7/1307.

Shang, C., Fu, L., Bao, X., Xu, X., Zhang, Y., and Xiao, H. (2022). Energy optimal dispatching of ship's integrated power system based on deep reinforcement learning. *Electric Power Systems Research*, 208, 107885.

Sommer, D.E., Yeremi, M., Son, J., Corbin, J.C., Gagné, S., Lobo, P., Miller, J.W., and Kirchen, P. (2019). Characterization and reduction of in-use ch4 emissions from a dual fuel marine engine using wavelength modulation spectroscopy. *Environmental Science & Technology*, 53(5), 2892–2899. doi:10.1021/acs.est.8b04244. URL https://doi.org/10.1021/acs.est.8b04244. PMID: 30712340.

Stamatakis, M.E. and Ioannides, M.G. (2021). State transitions logical design for hybrid energy generation with renewable energy sources in lng ship. *Energies*, 14(22). doi:10.3390/en14227803. URL https://www.mdpi.com/1996-1073/14/22/7803.

Sutton, R.S. and Barto, A.G. (2018). *Reinforcement learning: An introduction*. MIT press.

Wu, P., Partridge, J., Anderlini, E., Liu, Y., and Bucknall, R. (2021a). An intelligent energy management framework for hybrid-electric propulsion systems using deep reinforcement learning.

Wu, P., Partridge, J., Anderlini, E., Liu, Y., and Bucknall, R. (2021b). Near-optimal energy management for plug-in hybrid fuel cell and battery propulsion using deep reinforcement learning. *International Journal of Hydrogen Energy*, 46(80), 40022–40040.

Wu, P., Partridge, J., and Bucknall, R. (2020). Cost-effective reinforcement learning energy management for plug-in hybrid fuel cell and battery ships. *Applied Energy*, 275, 115258.