

Enhancing Transient Performance of Propulsion Engines in Actual Sea Using Reinforcement Learning-Based Proportional-Integral-Derivative Gain Tuning

Oleksiy Bondarenko¹, Yasushi Kitagawa², Ryohei Sawada³ and Tetsugo Fukuda¹

Received: 07 February 2025 / Accepted: 08 May 2025

© Harbin Engineering University and Springer-Verlag GmbH Germany, part of Springer Nature 2026

Abstract

The engine speed control system is critical to marine propulsion, particularly for ships navigating in wave conditions. Fluctuations in propeller torque, caused by wave-induced variations in the ship's speed and direction, can lead to undesirable effects such as increased fuel consumption and greenhouse gas emissions, increased engine wear and tear, and degraded performance. In the field of marine diesel engines, a proportional–integral–derivative (PID) control algorithm is the most widely used method for speed regulation. Aiming to maintain a constant engine speed, the PID governor continuously adjusts the fuel injection into the engine cylinders in response to propeller torque fluctuations. The quality of control, and thus engine performance, depends on a set of PID control gains. However, these optimal PID gains are typically determined during engine testbed commissioning and remain fixed, which leads to suboptimal performance under varying operational conditions. While model-based tuning techniques can achieve satisfactory results in relatively simple control tasks, particularly when the number of tunable parameters is small, adaptive tuning of controller parameters, while meeting multiple, often conflicting objectives, presents a considerable challenge. This paper explores the application of a reinforcement learning (RL) algorithm to adaptively tune PID gains, aiming to optimize engine performance under varying sea conditions and operational demands. A specific reward function is designed to effectively balance variations in engine shaft rotational speed and fuel pump index, ensuring optimal engine efficiency. The performance of the trust-region policy optimization and proximal policy optimization algorithms is evaluated for their suitability in continuous-action space control, demonstrating stable convergence and robustness. Simulation results indicate that the proposed RL algorithm outperforms the reference PID controller, offering promising results for adaptive gain scheduling in propulsion system control.

Keywords Propulsion engine; Proportional–integral–derivative speed control; Adaptive tuning; Reinforcement learning

1 Introduction

Maritime transportation is recognized as a notable contributor to environmental pollution, accounting for up to 3% of global greenhouse gas (GHG) emissions (IMO, 2020). In response, the International Maritime Organization (IMO) has introduced various regulations to reduce GHG emis-

sions from ships. One of the key regulations is the Energy Efficiency Design Index, which establishes energy efficiency standards for newly constructed ships. Simultaneously, the Energy Efficiency Operational Indicator, along with the Ship Energy Efficiency Management Plan, focuses on monitoring and improving the energy efficiency of ships in operation. However, the operation of ships involves a wide range of sea states, and insufficient attention has been given to the fuel efficiency of propulsion engines, which is a major contributor to GHG emissions under realistic operating conditions. Two components affect the performance of the propulsion engine under wave conditions: the average increase in engine load and the transient response, both of which are due to the propeller torque. While the first component can be managed by measures such as reducing the ship's speed and course, the engine speed control system can affect the second component. The engine speed control system, which uses a PID governor, continuously adjusts the fuel injected into the engine cylinders in response to fluctuations in propeller torque. Thus, the greater the fluctuation, the more substantial the transient response of the engine, consequently leading to reduced engine fuel

Article Highlights

- RL-based gains adapt to changing sea conditions.
- The system improves efficiency by balancing responsiveness and stability.
- Simulations show RL-tuning outperforms fixed-gain PID control.

✉ Oleksiy Bondarenko
bond@m.mpat.go.jp

¹ Power and Energy System, National Maritime Research Institute, Tokyo 1810004, Japan

² Ship Hydrodynamics Performance Evaluation, National Maritime Research Institute, Tokyo 1810004, Japan

³ Knowledge & Data System, National Maritime Research Institute, Tokyo 1810004, Japan

efficiency. This relationship between main engine load fluctuation and fuel consumption under wave conditions was confirmed by research conducted by Iwano et al. (2009). This research focused on developing a new engine shaft rotational speed control designed to operate the engine efficiently in wave conditions. The control objective was set to maintain a constant amount of fuel input, thereby providing a basic level of speed management and ensuring that the engine operates in a high fuel efficiency range. A test conducted on a container vessel showed a 5% to 6% improvement in efficiency, with an average of 1.3% improvement under severe sea conditions. Analysis of time series data also showed that this control system stabilized the thermal state of the propulsion engine. Therefore, modern engine shaft rotational speed governors currently operate in two discrete modes: a high-gain mode for speed stability in rough sea conditions and a low-gain mode to maintain a constant fuel input in mostly calm sea conditions. Consequently, adaptive tuning of the engine speed control system, depending on the sea state, is expected to reduce fluctuations in both engine shaft rotational speed and fuel input. This approach not only ensures the safe operation of the main engine but also enhances the fuel efficiency of the propulsion engine.

Numerous tuning methods have been proposed and widely implemented to adjust PID gains. Notable examples include H_∞ control (Xiros, 2002), fuzzy algorithm (Xu et al., 2007), and nonlinear control (Zhou et al., 2011). These methods primarily aim to optimize performance by enhancing response speed and minimizing transition time. However, optimizing multiple conflicting objectives remains a challenging problem. Additionally, expert control methods, which combine expert knowledge with control theory, have been employed to restructure control algorithms and dynamically adjust PID gains (Karray et al., 2002). Despite their advantages, most of these approaches can be classified as gain scheduling, often resulting in reduced capabilities (Calvo Rolle et al., 2009). Additionally, self-tuning strategies based on expert control are limited by the level of expertise involved and struggle to address uncertainties or adapt effectively to diverse control scenarios.

In recent years, machine learning has made substantial contributions to process control, promoting the theory of achieving optimal system performance under uncertain conditions based on evaluative rather than instructive feedback. This approach is frequently referred to as reinforcement learning (RL), which has evolved from approximate dynamic programming (Bertsekas, 2012). RL comes in different forms, some of which may involve partial knowledge of the environment or the system, while others do not. In the basic RL framework, an agent observes the current state of the environment, selects an action from a set of available actions to interact with the environment, alters the current state, and then evaluates the effectiveness of its action

based on the resulting impact. The reward function acts as a quantitative measure of the agent's decision regarding which action to take for a specific state, directly aligning with the control objective. The algorithm then attempts to learn the optimal policy incorporated in the agent, which aims to achieve the control objective as efficiently as possible by exploring the space of potential states and actions. RL utilizes the high expressive power of the deep neural networks to approximate uncertain and dynamic environments. This enables the system to learn an optimal policy that maps observed states to actions to maximize cumulative reward. Consequently, the system can adapt and improve performance over time, enhancing overall control effectiveness.

Furthermore, advanced techniques in RL application to adaptive PID tuning have demonstrated substantial improvements in control accuracy and robustness under uncertain conditions. For example, a deep RL-based adaptive PID tuning strategy was proposed for nonlinear MIMO systems with time-varying uncertainties, notably enhancing setpoint tracking accuracy and outperforming conventional PID controllers in chemical engineering applications (Wang and Ricardez-Sandoval, 2024). Similarly, the advantage actor-critic (A2C) algorithm was successfully applied to robotic systems, highlighting the potential of deep RL to dynamically tune PID parameters based on task-specific conditions. This approach enabled adaptive and optimized robot performance (van Veldhuizen, 2020). An adaptive RL-based PI controller utilizing deep deterministic policy gradient (DDPG) was introduced for DC motor speed control, demonstrating superior stability and adaptability compared to classical methods (Alejandro-Sanjines et al., 2023). Additionally, a constrained RL approach was formulated to specifically address safety concerns and minimal wear during adaptive PID tuning in industrial processes. Practical experiments validated the approach, revealing its effective real-time adaptability under varying conditions (Dogru et al., 2022). Furthermore, a comparative study of various RL algorithms (A2C, PPO, SAC, and DDPG) for PID tuning in a ball-and-beam system demonstrated the practical viability and improved disturbance rejection capabilities of RL-tuned controllers (Blake and Aphiratsakun, 2024).

Building upon these insights, the current paper aims to explore the feasibility of an RL-based adaptive PID gain scheduling strategy specifically tailored for propulsion engine speed control in marine vessels under realistic sea conditions. This work introduces a specific reward function design, which plays a crucial role in guiding the agent's learning process and enhancing overall performance. Unlike conventional approaches, the reward function in this study explicitly incorporates responses of engine shaft rotational speed and fuel pump index to fluctuations in propeller torque, considering their respective contributions to variations in delivered engine power. This function employs a scaled Gaussian transformation to balance sensitivity and

ensure thorough exploration within the PID gain parameter space. The objective of this adaptive tuning approach is dynamic, depending on the vessel's operating conditions, and includes control modes for the engine torque constant and the engine shaft rotational speed constant. Trust-region policy optimization (TRPO) and proximal policy optimization (PPO) algorithms are employed to determine the optimal policy for scheduling the governor gains. The performance of the proposed dynamic tuning algorithm is compared to that of a fixed-gain PID controller. Simulation results demonstrate promising outcomes for controlling the propulsion system.

2 Simulation environment and system modeling

The core elements of the RL framework are the agent, reward function, and environment. The agent interacts with its environment continuously by selecting actions, A_t , for the state S_t at step t , and observing the response of the environment, S_{t+1} , with performance feedback as a reward, R_{t+1} . Over time, the agent attempts to maximize the rewards, reinforcing good actions while discouraging bad ones, ultimately learning an optimal policy. Performance improvement is achieved through a balance of exploiting the best actions and exploring a wide variety of action-state-reward combinations. Thus, RL requires numerous experiences for training, and simulations offer a quick and low-cost environment to fulfill this requirement. In the context of dynamic tuning PID gains, the RL environment is modeled after the target ship's propulsion system operating actual sea conditions. The core components of the ship's propulsion system comprise three key components: the hull, the shaft with the propeller, and the propulsion engine, with the speed control system integrated into the engine. Detailed models of these components were developed in previous studies (Bondarenko and Fukuda, 2020; Kitagawa et al., 2019; Kitagawa and Kashiwagi, 2019) and have been extensively validated under actual sea conditions (Bondarenko and Kitagawa, 2024). A brief overview of these models is provided below.

2.1 Hull-propeller model

The Panama-size bulk carrier serves as the model ship for the environment. Its hull form was designed at the National Maritime Research Institute (NMRI) with a target speed of 14.5 kt at 90% maximum continuous rating (MCR). Table 1 presents the principal data of the ship. An assumption is that a ship moves at a constant speed, U_0 , while experiencing periodic longitudinal (surge) motion in waves. Figure 1 depicts the coordinate system associated with the ship's hull. The dynamics of the longitudinal velocity are given by the following:

$$(m + m_x) \frac{du_s}{dt} = X_p - R_t - X_w \quad (1)$$

where m and m_x represent the constant hull mass and added mass, respectively; X_p is the effective propeller thrust, R_t is the calm water resistance, and X_w is the added resistance due to waves.

Table 1 Principal particulars of a ship

Length, L (m)	Breadth, B (m)	Draught, D (m)	Displacement, ∇ (ton)	Service speed, U_0 (kt)	Block coefficient, C_B
217.0	32.3	12.2	7.17e4	14.5	0.84

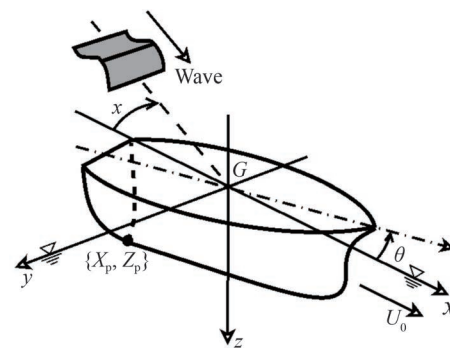


Figure 1 Body-fixed coordinate system of a ship advancing in waves

The effective propeller thrust and torque models are derived from the open-water characteristics of the propeller, which are obtained under calm water conditions. These models are expressed as follows:

$$\begin{aligned} X_p &= (1 - t_p) \rho n_p^2 D_p^4 K_t(J) \\ Q_p &= \rho n_p^2 D_p^5 K_q(J) \frac{1}{\eta_r}, \quad \therefore J = \frac{u_p}{n_p D_p} \end{aligned} \quad (2)$$

where t_p is the thrust deduction ratio, ρ is the water density, n_p is the propeller rotational speed (which, in the case of direct drive, is equal to the engine shaft rotational speed n_e), and D_p is the propeller diameter. K_q and K_t are the torque and thrust coefficients, respectively, expressed as polynomial functions of the J , which is defined by the open propeller characteristic. J represents the propeller advance ratio, η_r is the propeller relative rotation efficiency, and u_p is the effective inflow velocity. Figure 2 depicts the open-water characteristics of the subject propeller, and principal information is listed in Table 2.

The model of the inflow velocity, described below, comprises two components: one due to the hull surge motion, u_{p_m} , and one due to the wave orbital motion, u_{p_w} :

$$\begin{cases} u_p = u_{p_m} + u_{p_w} \\ u_{p_m} = (1 - w_p) \{ U_0 - \omega_e \zeta_a \sin(\omega_e t - \varepsilon_\zeta) \} \\ u_{p_w} = \alpha_w \omega \zeta_w e^{-kz_p} \cos \chi \cos(\omega_e t - kx_p \cos \chi) \end{cases} \quad (3)$$

where w_p is the propeller effective wake rate, ω_e is the encounter wave frequency, ζ_a and ε_ζ are the surge motion amplitude and phase, respectively; ω is the incident wave frequency, χ is the heading angle, ζ_w is the wave amplitude, k is the wavenumber ($= 2\pi/\lambda$), and λ is the wavelength, x_p and z_p are the coordinates of propeller position in terms of the hull center of gravity, G .

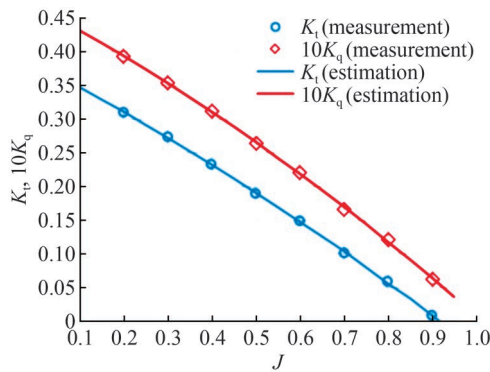


Figure 2 Propeller open-water characteristics

Table 2 Principal information of a propeller

Diameter, D_p (m)	Boss ratio	Number of blades	Moment of inertia, I_p (kg·m ²)	Added moment of inertia, $I_{p,add}$ (kg·m ²)
7.10	0.180	4	2.71e4	6.77e3

The inflow velocity component, u_{p_w} , due to wave orbital motion, is further modified by an effective wave amplitude coefficient α_w (Kitagawa et al., 2019). The calm water resistance, R_p , is evaluated in accordance with the ITTC recommended procedure.

$$\begin{cases} R_t = \frac{1}{2} \rho S_w U_0^2 C_T \\ C_T = (1 + \kappa) C_{F_0} + \Delta C_F + C_w \end{cases} \quad (4)$$

where R_t represents the calm water resistance (N); ρ is the fluid density (kg/m³); S_w is the wetted surface area (m²), C_T is the total resistance coefficient, κ is the form factor, C_{F_0} is the equivalent plate drag coefficient, ΔC_F is the roughness correction, and C_w is the wave resistance coefficient. The coefficient characteristics were analyzed based on the results of resistance tests conducted in the towing tank.

The component of added resistance in waves, introduced in Equation (1), is calculated using the added resis-

tance response amplitude operator (RAO):

$$X_w = \frac{4\rho g \zeta_w B^2}{L} K_{AW} \left(Fn, \frac{\lambda}{L}, \chi \right) \quad (5)$$

where g is the acceleration due to gravity and Fn is the Froude number. X_w is added resistance due to waves (N); ζ_w is the wave amplitude(m); λ is the wavelength (m); χ is the encounter wave angle (rad); K_{AW} is the added resistance response amplitude operator.

The K_{AW} RAO for a given hull speed and wave condition is obtained through interpolation over a set of pre-calculated data for different hull speeds, wave directions, and frequencies. For the target ship model, this database was generated using VESTA software (Tsujimoto et al., 2015), developed by the NMRI. Figure 3 illustrates the added resistance RAO for a set of speeds and head wave case ($\chi = 180^\circ$).

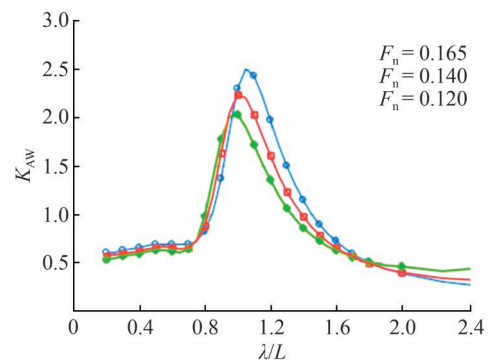


Figure 3 RAO of added resistance in regular head waves

Similarly, the amplitude and phase of the hull surge motion, which are required to evaluate the inflow velocity component, are determined through interpolation over a set of pre-calculated RAOs. These RAOs were computed using the strip method (Kitagawa and Kashiwagi, 2019), a technique recently developed by one of the authors. Figure 4 visualizes the RAO of surge motion components for a range of speeds and head wave cases.

Notably, the ship hull longitudinal velocity calculation procedure, as defined by Equation (1), focuses on the time evolution of the steady-state component (average) of the ship's speed. By contrast, the calculation of the propeller's effective thrust is based on the assumption that the effective inflow velocity changes with waves, as described in Equation (3). As a pragmatic approach, the time-averaged propeller effective thrust is considered, with the averaging period selected to correspond to the double encounter wave period.

2.2 Engine model

The shaft rotation dynamics interface the propeller with the engine through the balance of torques, as expressed below.

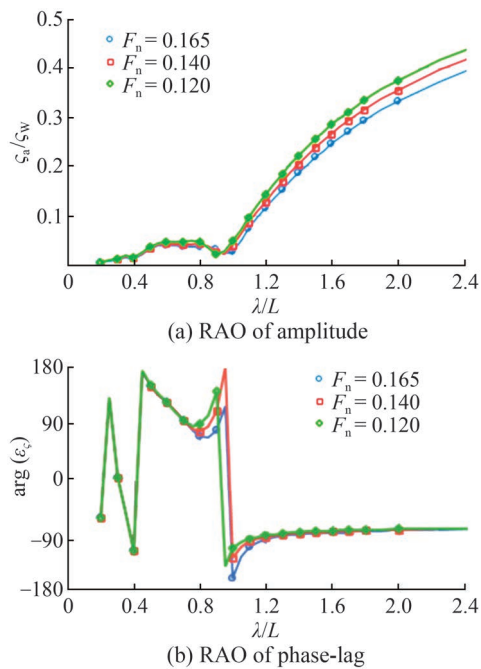


Figure 4 RAO of amplitude and phase-lag for surge motion in regular head waves

$$2\pi I_{sh} \frac{dn_e}{dt} = Q_e - Q_p \quad (6)$$

where Q_e and Q_p are the engine and propeller torques, respectively, I_{sh} is the moment of inertia of the rotating shaft system, including the engine, propeller, and added mass of water, and n_e is the engine shaft rotational speed.

In actual sea conditions, the fluctuating inflow velocity, as defined in Equation (3), leads to variations in propeller torque. In response, the engine control system adjusts the engine torque to maintain a constant shaft speed, which, in turn, affects the engine's performance. Thus, the task of the engine model is to capture the dynamics of engine torque, considering the engine efficiency under transient conditions.

The internal combustion engine is a complex system. In the field of propulsion systems simulation, a cycle mean value (CMV) engine modeling approach is commonly used for evaluating steady-state and transient performance (Hendrics, 1989; Theotokatos, 2010). In CMV modeling, the engine is represented as a series of control volumes connected through flow restrictions, where air and exhaust gas flow continuously, effectively ignoring the intermittent nature of the cylinder processes. The engine is decomposed into a finite number of elements, including the cylinder, air and exhaust gas receivers, turbocharger (TC) with a compressor and turbine, and an air cooler, as shown in Figure 5. Here Q_T is the turbine torque (N·m); Q_C is the compressor torque (N·m); h is specific enthalpy (J/kg), G is mass flow.

The compressor and turbine are mechanically linked via the TC shaft. An air cooler is positioned between the compressor and the air receiver. The fuel pump, that is, the fuel

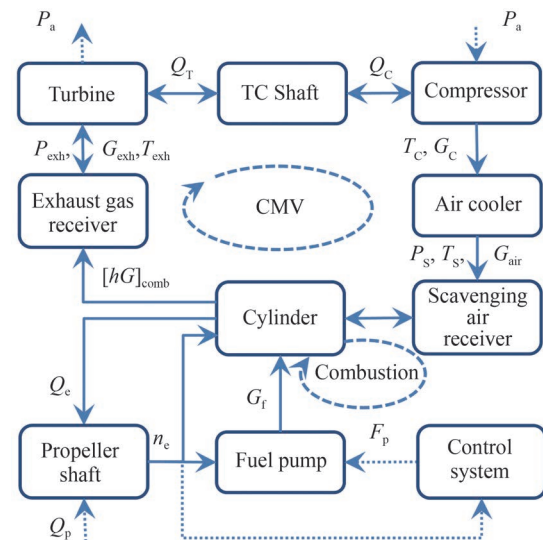


Figure 5 Composition of the propulsion engine

injector, is directly connected to the cylinder unit. The cylinder units of the engine are then connected to the propeller through shaft rotational dynamics, as expressed by Equation (6).

The engine torque is determined by the brake mean effective pressure, P_b , which is the difference between the indicated mean effective pressure (IMEP), P_i , developed in the cylinder volume V_s during one cycle and friction mean effective pressure, P_f . This relationship can be expressed as follows:

$$Q_e = \frac{P_b V_s}{2\pi}, \quad P_b = P_i - P_f \quad (7)$$

In turn, the developed IMEP is a function of various engine states, such as pressure (p), temperatures (T), and mass flows (G). The fundamental equations governing the temporal evolution of the engine's state variables are derived from the conservation laws of mass and energy, combined with the ideal gas equation.

$$\frac{dm}{dt} = \sum_i G_i \quad (8)$$

$$P = \frac{m}{V} \tilde{R}T \Rightarrow \frac{dP}{dt} = \frac{dm}{dt} \frac{\tilde{R}T}{V}, \quad \because T = \text{const} \quad (9)$$

$$c_v m \frac{dT}{dt} + u \frac{dm}{dt} = \sum_i [hG]_i \quad (10)$$

where u is the internal energy, $\tilde{R}$ is the specific gas constant, c_v is the specific heat capacity (isochoric), and m is the mass of gas (kg), T is the temperature (K); V denotes the volume of the considered element, P is the pressure, and the subscript i indicates the i th species entering or leaving the element.

The air and exhaust gas mass flow rates, through the engine and turbine, respectively, can be calculated under the assumption that an orifice with an equivalent mean effec-

tive flow area, $\mu\tilde{A}$, can characterize the engine cylinder as well as the turbine. Thus, the flow of compressible gas is evaluated in accordance with the following:

$$G_{(\text{air}, \text{exh})} = \mu\tilde{A} \frac{P_{\text{in}}}{\sqrt{\tilde{R}T_{\text{in}}}} \Psi(P_{\text{in}}, P_{\text{out}}) \quad (11)$$

where the subscripts, in/out, represent the inlet and outlet parameters of the considered element, respectively. $\Psi = f(P_{\text{in}}, P_{\text{out}})$ is the throttling characteristic of the orifice.

The TC, as an integral part of the engine, plays a crucial role in maintaining the sustainable airflow necessary for efficient fuel combustion. The dynamics of the TC, in relation to the rotational speed, are derived using the conservation of angular momentum, as expressed in the following form:

$$2\pi I_{\text{tc}} \frac{dn_{\text{tc}}}{dt} = Q_{\text{T}} - Q_{\text{C}} \quad (12)$$

where Q_{T} is the torque generated by the turbine from the expansion of exhaust gas, Q_{C} is the torque required by the compressor to compress air and deliver it to the scavenging receiver, n_{tc} is the rotational speed of the TC, and I_{tc} is the inertia of the TC shaft.

Considering the isentropic process of air compression and gas expansion, the torque required by the compressor and the torque delivered by the turbine are defined as follows:

$$Q_{\text{C}} = \frac{G_{\text{air}} C_{p,a} T_a}{\eta_{\text{ic}}} \left[\left(\frac{P_s}{P_a} \right)^{\frac{\gamma_{\text{air}} - 1}{\gamma_{\text{air}}}} - 1 \right],$$

$$Q_{\text{T}} = G_{\text{exh}} C_{p,e} T_{\text{exh}} \eta_{\text{it}} \left[1 - \left(\frac{P_a}{P_{\text{exh}}} \right)^{\frac{\gamma_{\text{exh}} - 1}{\gamma_{\text{exh}}}} \right] \quad (13)$$

For the compressor matched with the engine, a set of operating points, under steady-state conditions, is arranged along a single curve on the compressor map. Thus, the compressor's isentropic efficiency, η_{ic} , can be considered constant and is defined at the MCR point (Xiros, 2002). In contrast, the isentropic efficiency of the turbine, η_{it} , is strongly influenced by the turbine's blade speed ratio. The following empirical expression can be adopted (Ray, 1980):

$$\eta_{\text{it}} = \eta_{\text{it}_0} \left[\text{BSR} (c_{T_1} - \text{BSR}) + c_{T_2} \right] \quad (14)$$

The energy flow rate, $[hG]_{\text{comb}}$, exiting the engine cylinder and contributing to the energy balance of the exhaust gas receiver in Equation (10), is calculated by considering the energy conservation within the cylinder, averaged over one engine cycle, as follows:

$$[hG]_{\text{comb}} = G_{\text{air}} C_{p,a} T_s + G_{\text{f}} H_{\text{U}} - W_{\text{i}} \quad (15)$$

where $C_{p,a}$ is the Specific heat capacity (isobaric) of air (J/(kg·K)). The engine fuel mass flow rate, G_{f} , is calculated

as a linear function of the fuel pump index, F_{p} , as follows:

$$G_{\text{f}} = Z_{\text{c}} m_{\text{fc}_0} F_{\text{p}} n_{\text{e}} \quad (16)$$

where m_{fc_0} is the mass of fuel injected per cycle at MCR, Z_{c} is the number of engine cylinders, and F_{p} is the fuel pump index determined by the engine control system.

Finally, IMEP is directly obtained from the cylinder's indicated work, W_{i} , as follows:

$$P_{\text{i}} = \frac{\eta_{\text{c}} W_{\text{i}}}{V_{\text{s}}}, \quad W_{\text{i}} \propto F_{\text{p}} \quad (17)$$

In the CMV approach, the relationship between the indicated work and the fuel pump index is simplified and considered proportional. However, in reality, the indicated work is determined through engine combustion simulation. By combining discrete combustion cycle simulations with the continuous dynamics of the CMV model, referred to as combined CMV (Bondarenko and Fukuda, 2020), a highly accurate estimation of engine performance and an improved evaluation of propulsion system efficiency under transient conditions can be achieved.

Notably, the indicated work in Equation (17) is modified by introducing the coefficient η_{c} , which accounts for the losses due to incomplete fuel combustion. In reality, the combustion process in an engine is highly susceptible to variations in the air-to-fuel ratio (AFR). Under transient conditions, the fuel amount, determined by the fuel index, F_{p} , is adjusted sufficiently in response to the engine's torque demand to maintain a constant engine shaft rotational speed, as shown in Equation (6). However, the air mass flow may lag behind the fuel supply due to the large inertia of the TC. Thus, under transient conditions, the optimal AFR is disrupted due to the delay in air supply, leading to engine performance degradation. This phenomenon must be accounted for in the dynamic engine model. A series of theoretical and experimental studies (Bondarenko and Fukuda, 2018) has revealed the characteristics of combustion process degradation due to the depletion of air charge in the cylinder. Thus, the following combustion efficiency coefficient is introduced into the model (Equation (17)):

$$\eta_{\text{c}} = f(\text{AFR}) = C_0 \text{AFR}^{\frac{b_0}{\text{AFR}}}, \quad \therefore \text{AFR} = \frac{G_{\text{air}}}{G_{\text{f}}} \quad (18)$$

where C_0 and b_0 are the coefficients that represent the combustion efficiency characteristics. Figure 6 illustrates a family of characteristic curves showing the relative change in the AFR for different coefficient selections. As depicted in the figure, combustion efficiency drops rapidly when the AFR falls below its nominal value ($\text{AFR} < 1.0$).

The principal particulars and parameters of the accepted engine are listed in Table 3.

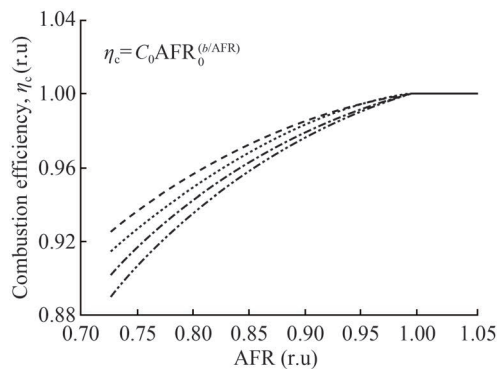


Figure 6 Combustion efficiency characteristics

2.3 Engine speed control

Propulsion engines are susceptible to fluctuations in propeller load, mainly due to the narrow optimal AFR bounds. Therefore, the control systems for marine diesel engines are designed to ensure stable speed and load operating regimes. Inadequate control system dynamics may lead to incomplete fuel combustion, thermal and mechanical overload of engine components, and excessive oscillations in shaft rotational speed. Typically, the propulsion engine control system includes a speed governor along with several auxiliary modules that provide protection against torque and thermal overload conditions.

Recently, digital governors that use a proportional-integral-derivative (PID) control law have become widely used for regulating the rotational speed of propulsion engines. PID controllers are effective in ensuring precise setpoint tracking, isochronous operation, and fast dynamic response. As indicated by its name, the PID control law generates a control action based on the values of the engine shaft rotational speed tracking error $e(t)$, its time integral, and possibly its time derivative, as expressed by the following relation:

$$u(t) = K_p e(t) + K_i \int_{t_0}^t e(\tau) d\tau + K_d \frac{d}{dt} e(t) \quad (19)$$

where K_p , K_i , and K_d represent proportional, integral, and differential gains, respectively. Different combinations of these PID terms produce a variety of control characteristics, which are generally bounded by two distinct modes: the speed-constant mode (represented by a vertical characteristic line on the engine's characteristic plane) and the torque-constant mode (characterized by deactivated governor action). Thus, adaptive tuning of these PID gains enables

optimal controller performance tailored to specific operational objectives of the propulsion system.

Another important consideration is the practical need for incorporating the derivative (D) term, which helps mitigate undesired oscillations in the controlled output, typically induced by the integral (I) term. However, the differentiator acts as a high-pass filter, allowing the propagation of high-frequency measurement noise through the closed-loop control system, often discouraging its practical implementation. Thus, advanced filtering techniques can substantially reduce noise in engine shaft rotational speed measurements. Moreover, the derivative of shaft rotational speed can be effectively estimated without directly differentiating the measured shaft speed signal itself (Xiros, 2002).

2.3.1 Torque limiter function

The PID control law is the primary linear component at the core of the engine governor. The practical implementation of PID for engine shaft rotational speed regulation requires several functional modules that ensure safe operation and remote control capabilities.

Thus, the output of the PID control logic, $u(t)$, interfaces with the engine's fuel pump via a servomotor, whose dynamic behavior can be described by the following:

$$T_{s,m} \frac{d}{dt} F_p(t) + F_p(t) = K_{s,m} u(t) \quad (20)$$

where $T_{s,m}$ and $K_{s,m}$ are the time constant and gain coefficient of the servomotor, respectively. These parameters are considered constant and thus remain outside the scope of adaptive tuning.

An additional functional module, known as the torque limiter function (TLF), ensures the safe operation of the power plant by preventing thermal and mechanical overload conditions. The TLF restricts the maximum allowable fuel flow rate, G_p , based on the current engine rotational speed by saturating the fuel pump index, F_p . An illustrative example of the TLF characteristic is provided in Figure 7.

During operation, the distance ΔF_p (or "gap") between the current mean value of the fuel pump index and the saturation boundary set by the TLF is critically important. This gap directly influences the available range of control actions during transient disturbances. A smaller gap increases the likelihood of reaching the saturation limit during rapid load changes, consequently restricting the controller's capability to maintain the desired engine performance. Thus, explicitly considering the available gap when evaluating and tuning PID gains through RL is crucial.

Table 3 Principal particulars and parameters of the engine WinGD W6X52

Nominal power (kW)	Nominal speed (r/min)	Specific fuel consumption (g/kW·h)	Engine moment of inertia I_e (kg·m ²)	Turbocharger moment of inertia I_{tc} (kg·m ²)	Mean effective pressure (kPa)	Cylinder volume (m ³)	Nominal air mass flow, (m ³ /s)	Nominal scavenge air pressure (kPa)	Compress or isentropic efficiency	Turbine isentropic efficiency
9065	88	168.0	5.8e4	4.02	2100	0.4916	19.5	360	0.81	0.75

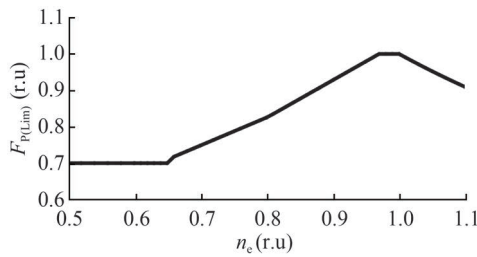


Figure 7 Fuel pump index limit characteristic

2.3.2 Admissible PID gain ranges and system stability

The dynamic response of the engine controlled by the PID speed governor is highly dependent on the chosen gain values. Deviations from the optimal gains can substantially degrade the system's transient performance. While the propulsion system, comprising an engine and propeller, is generally inherently stable due to the negative feedback from the propeller load, adaptive gain scheduling could potentially push the system toward its stability boundaries. Consequently, defining and analyzing the admissible range of PID gains for adaptive tuning is of considerable importance.

In marine control practice, a common approach for system tuning and stability assessment involves the use of transfer functions, which relate the response of a state variable of interest to input disturbances. The major state variables of a direct-drive turbocharged marine engine are the rotational speeds of the engine-propeller shaft and the TC shaft. These variables are governed by the differential equations presented in Equations (6) and (12). The engine and TC torques, as integral components of these equations, are inherently complex and nonlinear. However, they can be represented by explicit, smooth mapping functions of the state variables and control input, specifically the fuel pump index (Xiros, 2004; Bondarenko and Kashiwagi, 2012). These mappings can be generated from the detailed numerical engine model, as briefly introduced in Section 2.2. A linearized model of the propulsion system can be derived through standard perturbation analysis, applying Taylor series expansions to the torque mapping functions. Without delving into the full derivation, the resulting closed-loop transfer function, which characterizes the engine shaft rotational speed response to a disturbance in propeller inflow velocity in the Laplace domain, can be expressed as follows:

$$W(s) = \frac{B_1(s)}{p_c(s)} \quad (21)$$

The closed-loop characteristic polynomial, $p_c(s)$, is given explicitly as follows:

$$p_c(s) = a_0 s^3 + a_1 s^2 + a_2 s + a_3 \quad (22)$$

where the coefficients a_0 , a_1 , a_2 , and a_3 directly capture the

PID control gains.

Evaluating the stability of systems characterized by a third-order characteristic polynomial is conveniently performed using the Vyshnegradsky diagram (Syromyatnikov, 1983).

Transforming Equation (22) into normalized form and introducing a new variable: $q = s \sqrt[3]{\frac{a_0}{a_3}}$, which yields the following:

$$q^3 + Aq^2 + Bq + 1 = 0 \quad (23)$$

where A and B are Vyshnegradsky parameters, which provide a convenient visual framework on the Vyshnegradsky diagram to distinguish between stable and unstable operating regions, and $A = a_1 \sqrt[3]{\frac{a_3}{a_0^2}}$, $B = a_2 \sqrt[3]{\frac{a_3}{a_0^2}}$.

The parameters A and B explicitly depend on the PID gains. Thus, adjusting these gains facilitates the direct plotting of the closed-loop system characteristics onto the Vyshnegradsky diagram to visually assess stability. Figure 8 illustrates the system's characteristics for varying values of K_p and K_i while maintaining K_d fixed at a minimum, whereas Figure 9 shows the same gain range with K_d fixed at a maximum. Both diagrams clearly indicate that the system's operating points are well within the stable boundaries, notably away from the stability margins, thus confirming that the adaptive PID gains scheduling provides a sufficient buffer against instability.

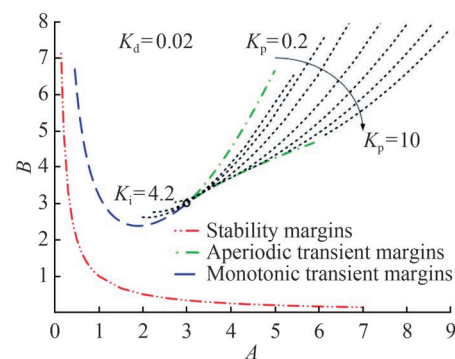


Figure 8 Closed-loop system characteristics at "min" K_d

3 Implementation of the RL framework

The RL algorithm seeks to discover the optimal policy for selecting control actions, $u_t = \pi^*(s_t)$, that maximizes the reward gathered over time. The policy π^* determines the decision-making capability of the agent, representing the probability that the action $A_t = a'$ will be chosen when the system is in state $S_t = s'$. The agent and the environment interact with each other, generating experiences comprising a set of state–action–reward pairs, which the agent

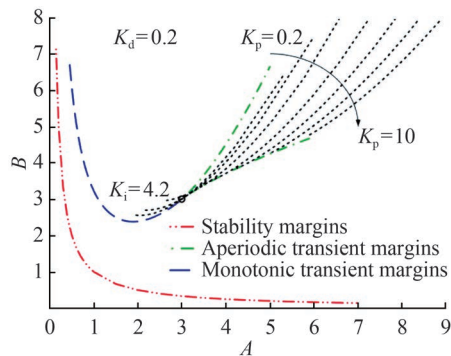


Figure 9 Closed-loop system characteristics at “max” K_d

uses to improve the policy. A wide variety of algorithms is available for learning the optimal policy. However, deep RL (DRL) algorithms, combining the ideas of deep learning and RL, have recently gained increasing popularity. The differences in these DRL algorithms depend on how the agent learns and selects actions. In the field of DRL, TRPO (Schulman et al., 2015) and PPO (Schulman et al., 2017) are two popular techniques within the family of actor–critic algorithms. TRPO works by having the critic estimate the state-value function, while the actor updates the policy based on trust-region constraints, ensuring learning stability. By contrast, PPO also incorporates an actor–critic approach similar to TRPO but differs by aiming to balance learning stability with sample efficiency. This balance is achieved by clipping the probability ratio between the new and old policies within a specified range. TRPO and PPO algorithms can be applied to control problems with either continuous or discrete action and state spaces.

In the actor–critic algorithms, both the policy function, known as the actor, and the state-value function, known as the critic, are approximated and learned using ANNs. The actor and critic functions are denoted as $\pi^*(S; \theta_a)$ and $V(S; \theta_c)$, respectively, where θ_a and θ_c represent the vectors of the ANN parameters.

The final key component of RL implementation is the design of the reward function, which is crucial in the learning process because it provides the primary feedback signal to the RL agent. The reward function design is discussed in detail in the subsequent chapter.

3.1 General justification of the reward function

The reward function establishes the relationship between state–action pairs and the corresponding rewards. Designing an appropriate reward function is crucial in RL because it directs the learning behavior of the agent and directly influences its capability to achieve the desired control objectives.

For a ship cruising in waves, the fluctuating inflow velocity directly affects the behavior of the rotating shaft system and the engine’s operating point. The responses of the engine in waves are critical for multiple reasons. First,

the margin from the operating point trajectory and the limits of the engine’s operating envelope are crucial for avoiding thermal overloading. Second, the response of the operating point primarily affects combustion efficiency, and consequently, the engine’s thermal efficiency. The position of the engine operating point is characterized by the delivered power and the shaft rotational speed. Therefore, the objective of adaptive tuning for the speed control system parameters is to manage engine control in such a way that variations in output power and engine shaft rotational speed are minimized simultaneously.

The engine delivered power is derived as follows:

$$N_{e,d} = 2\pi n_e Q_e \quad (24)$$

The engine torque, Q_e , is generally considered proportional to the fuel pump index F_p , that is, $Q_e \propto F_p$, across most of its operating range. Thus, the engine’s delivered power can conceptually be understood as a function of control effort and controllable variables, such as the fuel pump index, F_p , and the engine shaft rotational speed, n_e . However, the nonlinearity due to multiplicative action in the power definition by Equation (24) restricts the reward of every parameter contribution. Aiming to address this issue, a linear representation that relates the relative change in output $N_{e,d}$ to the relative change in inputs n_e and F_p is derived in this study. This transformation can be achieved using the method of small increments. The approach involves decomposing the function by applying a logarithm, differentiating the function, and then substituting relative increments for differentials. The resulting transformation yields the following:

$$\begin{aligned} \log(N_{e,d}) &= \log(2\pi) + \log(n_e) + \log(Q_e) \\ &\Downarrow \\ \frac{dN_{e,d}}{N_{e,d}^0} &= \frac{dn_e}{n_e^0} + \frac{dQ_e}{Q_e^0} \\ &\Downarrow \\ \delta N_{e,d} &= \delta n_e + \delta F_p, \quad \because \frac{dQ_e}{Q_e^0} \equiv \frac{dF_p}{F_p^0} \end{aligned} \quad (25)$$

where the superscript “0” denotes the steady-state values of the corresponding variables.

As a result, small increments in delivered power are approximated as linear combinations of small deviations in the fuel pump index and engine shaft rotational speed. This approximation facilitates the explicit quantification of the engine’s response to inflow velocity disturbances. Specifically, relative deviations in relation to inflow velocity disturbances u_p are defined as follows:

$$\frac{\delta N_{e,d}}{\delta u_p} = \frac{\delta n_e}{\delta u_p} + \frac{\delta F_p}{\delta u_p}, \quad \because \delta u_p \equiv \frac{\Delta u_p}{U_p^0} \quad (26)$$

The developed environment for the hull-propeller-engine system can be used to determine the behavior of the contributing components to the delivered power deviation, as described by Equation (26), in response to sinusoidal inflow velocity disturbances and the effect of PID gain settings. Figure 10 visualizes explicitly the values of the components $\delta n_e / \delta u_p$ and $\delta F_p / \delta u_p$ for various proportional gain settings K_p and different positions of the engine operating point, determined by the shaft speed set point, n_{sp} .

3.2 Scaling the reward function

Although evaluating the sensitivity of the system to PID gain settings is feasible, the contributions of both components to the power response are not equivalent. As illustrated in Figure 10, the response of the fuel pump index is more substantial than that of the engine shaft rotational speed for the same disturbance. Therefore, to ensure balanced sensitivity and avoid skewed reward signals favoring particular parameters, a scaling function is introduced for each component that contributes to the reward. Based on empirical considerations from Figure 10, a Gaussian-type exponential scaling function is proposed to normalize the contributions and prevent dominance by any single parameter:

$$f(x) = \exp\left(-\frac{x^2}{\text{scale}}\right) \quad (27)$$

The parameter “scale” controls the sensitivity and balance, ensuring smooth and evenly weighted feedback to the RL agent, which helps in achieving stable convergence during training. Figure 11 depicts the results of the scaling function transformation.

Finally, the reward function can be represented as follows:

$$R = f_1\left(\frac{\delta F_p}{\delta u_p}\right) + f_2\left(\frac{\delta n_e}{\delta u_p}\right) \quad (28)$$

With the defined reward function, the RL agent aims to maximize the cumulative reward across various sea states, characterized by different amplitudes and frequencies of inflow velocity fluctuations. This condition is achieved by providing a tradeoff between the responses of the engine shaft rotational speed and the fuel pump index.

Moreover, practical implementation highlighted the need to introduce a penalty term to discourage the RL agent from repeatedly choosing PID gain values at the boundaries of admissible ranges. Aiming to prevent this phenomenon, a regularization term was added to the reward, as follows:

$$r = -\gamma \sum_{i=1}^3 |\text{action}_i|, \gamma \in (0, \dots, 1.0), \text{action} \rightarrow \{K_p, K_i, K_d\} \quad (29)$$

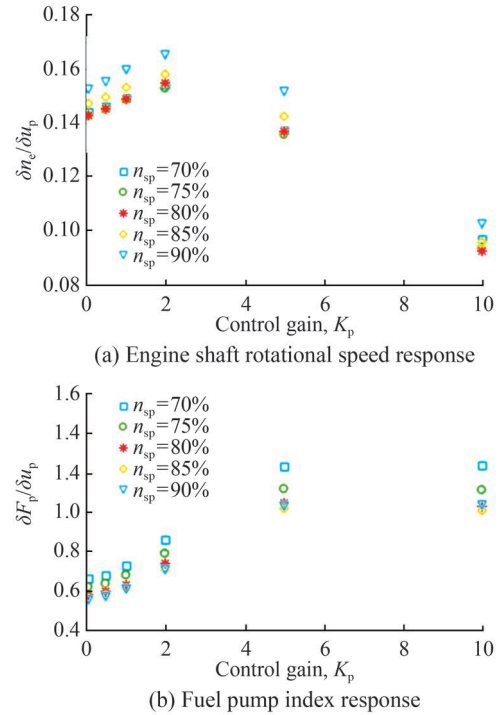


Figure 10 Response of engine shaft rotational speed and fuel pump index to a fluctuation of the inflow velocity

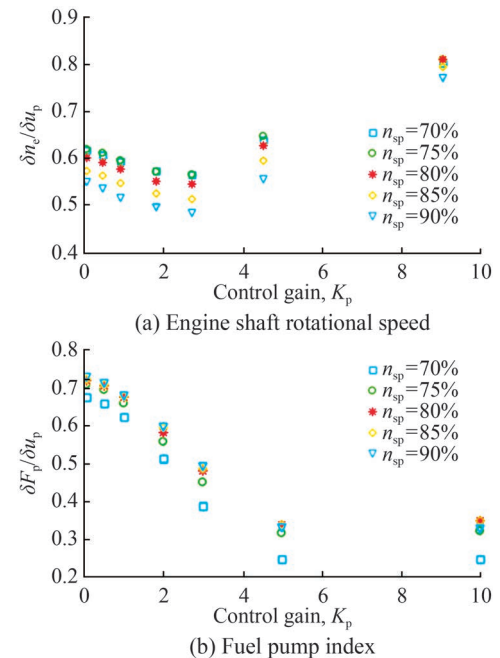


Figure 11 Scaled components of the reward function

4 Experiments, results, and discussion

To evaluate the proposed RL-based PID tuning framework a series of simulation experiments were conducted using MATLAB Simulink and the Reinforcement Learning Toolbox. This section describes the configuration of the simula-

tion environment, the results of the training process, and an analysis of the impact of the reward regularization term. Two representative RL algorithms, TRPO and PPO, are compared with a fixed-gain PID baseline to assess performance improvements in various operational conditions.

4.1 Experimental setup

The complete environment was implemented and simulated using MATLAB Simulink®, with the Reinforcement Learning Toolbox™ facilitating the interaction between the agent and the environment for training and validation. For the PPO and TRPO methods, the state-value function was implemented as a neural network comprising a single fully connected layer with 55 units and a rectified linear unit activation function. Two types of policy representations were employed: a stochastic actor for the continuous action space and a discrete categorical actor for the discrete action space. The continuous actor was implemented as a neural network with one fully connected layer of 55 units and a hyperbolic tangent activation function. However, the discrete actor comprised only an output layer, with a size equal to the number of discrete action combinations, and applied a SoftMax function at the output.

The state representation for the learning algorithm was defined using available propulsion system performance parameters, all of which were normalized relative to the maximum continuous rating (MCR) point. Given that the learning objective was to evaluate the controller's performance under harmonic disturbance, the state vector incorporated the estimated inflow velocity's variance and period (Bondarenko and Kitagawa, 2021), alongside the average value and variance of each measured parameter. In total, 18 state variables constitute the state vector, as listed in Table 4. Additionally, to indicate the available margins within the engine's operational envelope, the distance between the current fuel pump index and the TLF saturation line was included in the state vector.

Cruising ships experience a wide range of operational

conditions, characterized by varying ship speeds (and thus engine loads), as well as different wave severities and directions. These factors all affect the frequency and amplitude of inflow velocity. Therefore, to promote generalization and exploration of the policy during training, every four-hour episode was composed of a random sequence of wave conditions with varying amplitudes and periods. Based on the assumption that ocean waves remain probabilistically steady for approximately 30 min (Lloyd, 1998), each wave condition lasted for 40 min, including a 10 min transition phase between states. Furthermore, each episode was initiated randomly at a different position of the engine operating point. The RL agent evaluated the state and updated its action every 240 s. Table 5 summarizes the learning environment and the accepted range of PID gains. For the case of a discrete action space, a full factorial plan was constructed, resulting in 1331 discrete action combinations.

Finally, the computational cost is primarily associated with simulating the environment, rather than the training algorithm itself. The implemented propulsion system environment is computationally efficient and typically operates faster than real time (Bondarenko and Fukuda, 2020; Bondarenko and Kitagawa, 2024). Consequently, despite the training process requiring multiple episodes of relatively long duration, it still remains computationally justified and manageable. Moreover, once the RL agent is trained, its deployment is computationally lightweight. The policy is represented by a simple neural network, making real-time implementation feasible even on embedded systems or low-power hardware.

4.2 Learning performance and results

The learning process for each version of the algorithm was repeated multiple times under different initial conditions to verify the consistency and repeatability of the results. Figure 12 shows a representative example of learning progress, depicted as the average cumulative reward across training episodes for all tested algorithms. The subscript “d” indicates agents trained using a discrete action space.

The generalized performance outcomes from all learning sessions are summarized in Figure 13 and compared with the reference PID controller. As shown in the figure, all RL-based algorithms clearly outperform the fixed-gain PID baseline, validating the effectiveness of adaptive gain tuning strategies.

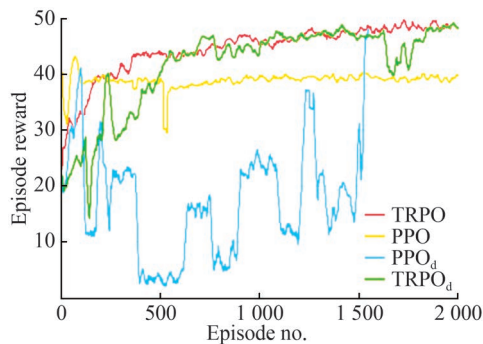
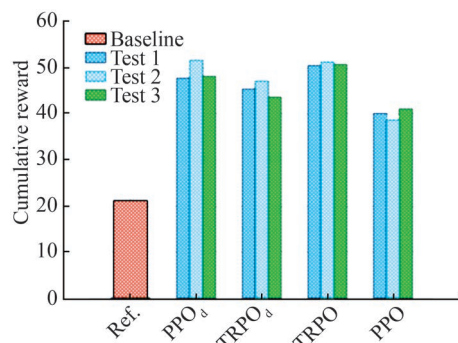
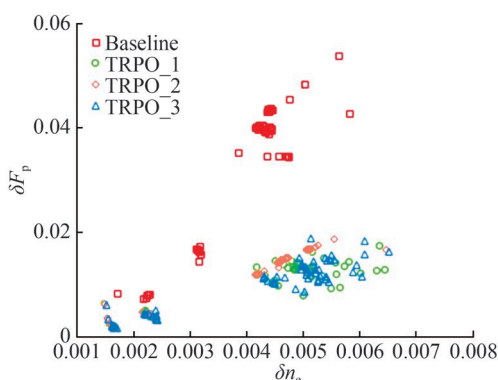
Further analysis of Figure 13 highlights that the TRPO-based agent consistently delivers stable results, outperforming the other algorithms. Aiming to explore this phenomenon in more detail, Figure 14 presents a parameter plane visualization of engine responses, focusing on the variances of the fuel pump index and engine shaft rotational speed. Each point corresponds to a time-domain evaluation performed at intervals specified for the RL environment (Table 5).

Table 4 List of variables in the state representation vector

Item	Derived metric
Engine shaft rotational speed, n_e	mean, variance
Turbocharger shaft rotational speed, n_{tc}	mean, variance
Speed set point, n_{sp}	—
Scav. air pressure, P_s	mean, variance
Exh. gas temperature, T_{exh}	mean, variance
Engine load	mean
Fuel consumption per transport work	mean
Ship speed, V_s	mean
Inflow velocity, u_p	mean, variance
Fuel pump index, F_p	mean, variance
TLF gap, ΔF_p	mean
Wave period, T_w	—

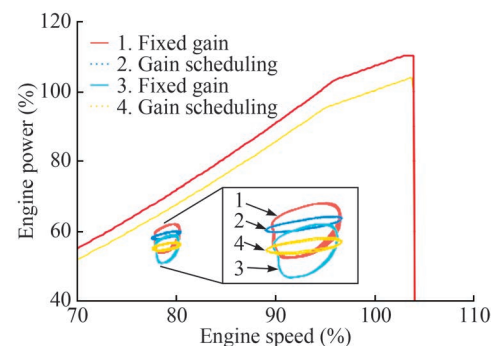
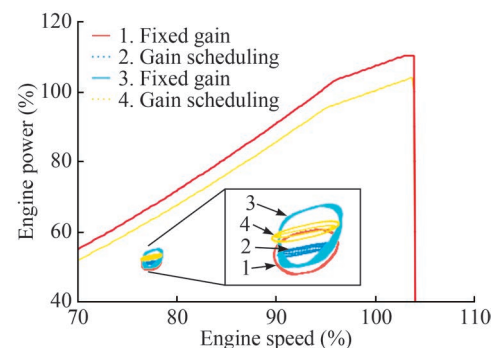
Table 5 Configuration of the environment

Wave amplitude (m)	Wavelength ratio	Speed set point, n_{sp} (%)	Proportional Gain, K_p	Integral gain, K_i	Differential gain, K_d	Episode duration (s)	Max episodes no.	Evaluation interval (s)	Baseline PID gains
0.2, ..., 1.2	0.4, ..., 1.2	76, ..., 90	0.5, ..., 10.5	0.2, ..., 4.2	0.02, ..., 0.2	14401	2000	240	{3.3 0.88 0.007}


Figure 12 RL's progress in terms of average cumulative reward

Figure 13 Generalized results of the agent learning

Figure 14 Parameter plane of the engine response

The results demonstrate that adaptive PID gain tuning using the TRPO algorithm effectively reduces fluctuations in the fuel pump index without increasing variations in engine shaft rotational speed, offering an improvement over the standard PID controller. This conclusion is further supported by the time-domain responses shown in Figures 15 and 16, which illustrate the evolution of the engine operating point across various operational scenarios. The RL-based

approach substantially reduces engine power fluctuations compared to fixed-gain PID control, reinforcing the advantages of adaptive tuning in dynamic marine propulsion environments.


Figure 15 Engine operating point evolution, scenario 1

Figure 16 Engine operating point evolution, scenario 2

Notably, while TRPO outperforms PPO, the primary objective of this study was not to conduct a detailed algorithmic comparison, but rather to assess the feasibility of applying RL to adaptive PID tuning in realistic marine propulsion conditions. Consequently, all algorithms were implemented with standard hyperparameter settings, without extensive tuning or optimization. The observed superiority of TRPO may be partly attributed to its trust-region policy update mechanism, which enforces stricter policy stability compared to PPO's clipped surrogate objective. This characteristic makes TRPO more suitable for sensitive, real-time control problems such as engine speed regulation, where abrupt policy changes can degrade system performance. By contrast, while PPO is generally more sample-efficient and widely adopted, it tends to be more sensitive to hyperparameter tuning. The relatively poor performance of PPO in this study likely reflects this sen-

sitivity, rather than any inherent performance limitation.

4.3 Effect of reward regularization strategy

The original regularization term, Equation (29), was incorporated into the reward function to prevent the agent from frequently selecting actions near the boundaries of the action space. However, its linear form, which proportionally increases as the action value approaches the boundaries, was found to be suboptimal. The results showed that the action space remained biased toward boundary values, indicating that the penalty for selecting extreme actions was insufficient.

Aiming to address this limitation, a U-shaped penalty function is introduced as the new regularization term. The new form is expressed as follows:

$$r = - \sum_{i=1}^3 \frac{2.0}{1.0 + \exp(-(|\text{action}_i| - 1)\omega)} \quad (30)$$

where the weight ω controls the steepness of the U-shape. As illustrated in Figure 17, this function penalizes actions aggressively only near the boundaries compared to the linear approach, thereby promoting more balanced exploration within the action space.

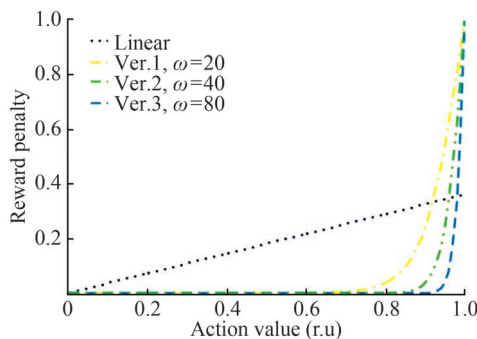


Figure 17 U-shaped penalty vs linear

Aiming to evaluate the effectiveness of the revised regularization term, one of the best-performing agents (TRPO) was selected and retrained using the new penalty function. Different values of the weight parameter $\omega = [20, 40, 80]$ were tested to examine their impact on the agent's behavior and performance. These correspond to Ver.1, Ver.2, and Ver.3, respectively, as shown in the figures.

Although the cumulative reward during training did not show a substantial difference compared to the original linear regularization, as shown in Figure 18, a qualitative improvement was observed in the action space analysis (Figure 19). The U-shaped penalty notably expanded the exploration space, demonstrating a more diverse selection of actions. This expanded exploration may improve learning efficiency and robustness, especially in more complex or less deterministic environments.

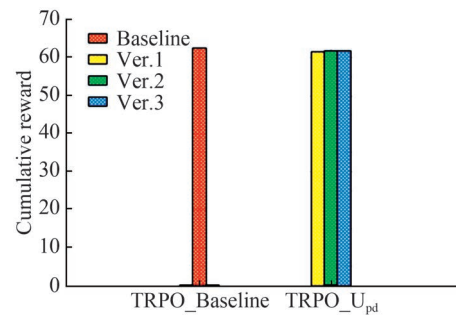


Figure 18 Performance of the updated TRPO agent with respect to the original

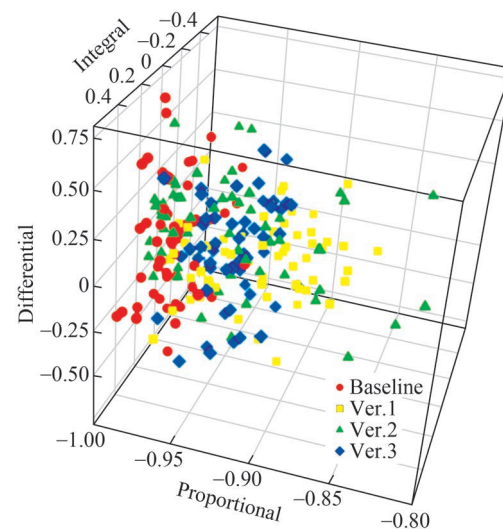


Figure 19 Agent's action exploration space

Although the selection of the optimal ω value depends on empirical testing, in general, larger ω values make penalties more transparent but may oversimplify the tradeoff between exploration and stability. A moderate $\omega = 40$ could be used as a default value.

The observations highlight the importance of designing effective regularization terms that align with the desired agent behavior and exploration strategy. Further research into dynamic or state-dependent weights for ω could improve these results and offer increased adaptability to varying environmental conditions.

5 Conclusions

The study addressed the challenge of engine speed control for marine propulsion systems operating under realistic sea conditions. By applying DRL for the adaptive tuning of PID controller gains, the approach demonstrated the capability to simultaneously optimize conflicting performance objectives, minimizing fluctuations in engine shaft rotational speed and variations in the fuel pump index. A

key factor in the success of this approach is the reward function design. A specially tailored reward function was introduced to balance exploration with effective convergence, ensuring that the adaptive controller's performance aligns with the desired engine performance. Stability analysis through Vyshnegradsky diagrams confirmed wide gain margins, ensuring that the adaptive PID gain scheduling operates safely within stable boundaries. Additionally, the introduction of a U-shaped penalty function notably improved the effectiveness of the RL algorithm by promoting comprehensive exploration within the gain space. Finally, TRPO and PPO algorithms were evaluated, with TRPO delivering more consistent and stable results than PPO, although the performance difference remained modest.

These findings highlight the feasibility and substantial advantages of RL-based adaptive tuning strategies in marine propulsion systems. However, practical validation on an actual vessel remains inherently challenging. Aiming to overcome this limitation, future research will employ a comprehensive hardware-in-the-loop validation approach developed in previous studies (Kitagawa et al., 2019; Bondarenko and Kitagawa, 2024). This innovative method, which combines a physical model ship and sea environment with numerical simulations of propulsion systems, will facilitate realistic and rigorous validation of the proposed adaptive PID gain scheduling algorithm, effectively bridging theoretical insights with practical maritime applications.

Nomenclature

Symbol	Description	Symbol	Description
C_T	Total resistance coefficient	K_d	Derivative gain of PID controller
C_{f0}	Viscous resistance coefficient of equivalent flat plate	K_i	Integral gain of PID controller
ΔC_f	Roughness correction	K_p	Proportional gain of PID controller
C_w	Wave resistance coefficient	K_q	Propeller torque coefficient
$C_{p,a}$	Specific heat capacity (isobaric) of air (J/(kg·K))	K_t	Propeller thrust coefficient
$C_{p,e}$	Specific heat capacity (isobaric) of exhaust gas (J/(kg·K))	$K_{s,m}$	Gain of the servomotor
D_p	Propeller diameter (m)	L_{pp}	Length between perpendiculars (m)
F_p	Fuel pump index	$P_{e,d}$	Engine delivered power (W)
G	Mass flow (kg/s)	P_b	Brake mean effective pressure (Pa)
G_f	Fuel mass flow rate (kg/s)	P_f	Friction mean effective pressure (Pa)
I_e	Engine shaft inertia (kg·m ²)	P_i	Indicated mean effective pressure (Pa)
I_{tc}	Turbocharger shaft inertia (kg·m ²)	Q_e	Engine torque (N·m)
I_p	Propeller shaft inertia (kg·m ²)	Q_p	Propeller torque (N·m)
$I_{p,add}$	Added moment of inertia for propeller (kg·m ²)	Q_T	Turbine torque (N·m)
I_{sh}	Inertia of the shaft system (kg·m ²)	Q_C	Compressor torque (N·m)
J	Advance ratio (propeller characteristic parameter)	$\check{R}$	Specific gas constant (J/(kg·K))
H_U	Fuel lower heating value (J/kg)	R_t	Calm water resistance (N)
hG	Energy rate (W)	S_w	Wetted surface area of hull (m ²)
Fn	Froude number	T	Temperature (K)
k	Wavenumber	T_w	Wave period (s)
		$T_{s,m}$	Time constant of the servomotor (s)
		t_p	Thrust deduction rate
		$e(t)$	Speed tracking error (r/min)
		m	Mass (kg)
		m_x	Added mass in the longitudinal direction of the hull (kg)
		m_{fc}	Mass of fuel per cycle (kg)
		n_e	Engine shaft rotational speed (r/min)
		n_{sp}	Engine speed set point (r/min)
		n_{tc}	Turbocharger rotational speed (r/min)
		P	Pressure (Pa)
		$u(t)$	Control signal from PID controller
		u_p	Effective inflow velocity to the propeller (m/s)
		$u_{p,m}$	Inflow velocity component due to hull motion (m/s)
		$u_{p,w}$	Inflow velocity component due to wave orbital motion (m/s)
		w_p	Effective wake rate of propeller
		W_i	Indicated work (J)
		X_p	Effective propeller thrust (N)
		X_w	Added resistance due to waves (N)
		V_s	Cylinder swept volume (m ³)
		Ψ	Throttling function
		α_w	Effective wave amplitude coefficient
		κ	Form factor
		ζ_w	Wave amplitude (m)
		ρ	Fluid density (kg/m ³)
		η_r	Propeller relative rotation efficiency
		η_i	Isentropic efficiency
		η_c	Combustion efficiency
		λ	Wavelength (m)

γ	Specific capacities ratio
ξ_a	Surge motion amplitude (m)
$\mu\tilde{A}$	Mean effective flow area (m ²)
ω_e	Encounter wave frequency (rad/s)
χ	Encounter wave angle (rad)

Subscripts

Symbol	Description
a	Ambient
air	Air
comb	Combustion
C	Compressor
T	Turbine
exh	Exhaust
0	Variable value at MCR
s	Scavenge

Competing interest The authors have no competing interests to declare that are relevant to the content of this article.

References

- Alejandro Sanjines U, Maisincho Jivaja A, Asanza V, Lorente Leyva LL, Peluffo Ordóñez DH (2023) Adaptive PI controller based on a reinforcement learning algorithm for speed control of a DC motor. *Biomimetics* 8(5): 434. <https://doi.org/10.3390/biomimetics8050434>
- Bertsekas DP (2012) *Dynamic Programming and Optimal Control*. Vol. II, 4th Edition: Approximate Dynamic Programming 2012. Athena Scientific, 4th ed
- Blake XJ, Aphiratsakun N (2024) Reinforcement learning for PID gain tuning in ball-and-beam systems: A comparative study. 2024 1st International Conference on Robotics, Engineering, Science, and Technology (RESTCON) 187-190. <https://doi.org/10.1109/RESTCON60981.2024.10463594>
- Bondarenko O, Kashiwagi M (2012) Ensuring safe operation of ship propulsion plant in extreme sea condition. 11th International Marine Design Conference (IMDC) 3: 83-96. <https://doi.org/10.13140/RG.2.2.30079.73124>
- Bondarenko O, Fukuda T (2018) Intelligent air management for ultimate flexibility of ship propulsion plant operation-1st Report: Experiment and Simulation. The Japan Institute of Marine Engineers 53(3): 348-354. <https://doi.org/10.5988/jime.53.348>
- Bondarenko O, Fukuda T (2020) Development of a diesel engine's digital twin for predicting propulsion system dynamics. *Energy* 196: 117126. <https://doi.org/10.1016/j.energy.2020.117126>
- Bondarenko O, Kitagawa Y (2021) Application of the Kalman filtering technique for nonlinear state estimation in propulsion system. *Journal of Marine Science and Technology* 26: 618-631. <https://doi.org/10.1007/s00773-020-00763-0>
- Bondarenko O, Kitagawa Y (2024) A comprehensive propulsion system analysis: an integrated approach utilizing engine simulation and free-running model experiments. *J. of Mar. Eng. & Techn.*, 1-12. <https://doi.org/10.1080/20464177.2024.2412442>
- Calvo Rolle JL; García RF, López García H, Machón González JJ (2009) Developed an expert system of an empirical method to choose correct expressions for PID controllers tuning in open loop. Proceedings of the 35th Annual Conference of IEEE Industrial Electronics. Piscataway: IEEE 1-6: 2044-2049. <http://dx.doi.org/10.1109/IECON.2009.5415064>
- Dogru O, Velswamy K, Ibrahim F, Wu Y, Sundaramoorthy AS, Huang B, Xu S, Nixon M, Bell N (2022) Reinforcement learning approach to autonomous PID tuning. *Computers & Chemical Engineering* 161: 107760. <https://doi.org/10.1016/j.compchemeng.2022.107760>
- Hendrics E (1989) Mean value modelling of large turbocharged two-stroke diesel engines. *Journal of Engines* 98: 986-998. <https://doi.org/10.4271/890564>
- IMO (2020) Fourth Greenhouse Gas Study. [accessed 2023 Nov 1]. <https://www.imo.org/en/ourwork/Environment/Pages/Fourth-IMO-Greenhouse-Gas-Study-2020.aspx>
- Iwano J, Hirooka H, Kondo K, Shigaki T (2009) Onboard testing results for new fuel efficient main engine governor system—Development of NABTESCO 70X. *Journal of Marine Engineering* 44(5): 105-110. <https://doi.org/10.5988/jime.44.787>. (in Japanese)
- Karray F, Gueaieb W, Al-Sharhan S (2002) The hierarchical expert tuning of PID controllers using tools of soft computing. *IEEE Transactions on Systems Man and Cybernetics, Part B-Cybernetics* 32(1): 77-90. <https://doi.org/10.1109/3477.979962>
- Kitagawa Y, Bondarenko O, Tsukada Y (2019) An experimental method to identify a component of wave orbital motion in propeller effective inflow velocity and its effects on load fluctuations of a ship main engines in waves. *Applied Ocean Research* 92: 101922. <https://doi.org/10.1016/j.apor.2019.101922>
- Kitagawa Y, Kashiwagi M (2019) A strip method including nx-related terms and its effects on propeller inflow velocity in waves. *Journal of the Japan Society of Naval Architects and Ocean Engineers* 29: 39-50. <https://doi.org/10.2534/jjasnaoe.29.39>
- Lloyd ARJM (1998) *Seakeeping, Ship Behaviour in Rough Weather*. Hoboken: John Wiley & Sons
- Ray A (1980) Dynamic modelling of power plant turbines for controller design. *Applied Mathematical Modelling* 4(2): 109-112. [https://doi.org/10.1016/0307-904X\(80\)90114-6](https://doi.org/10.1016/0307-904X(80)90114-6)
- Schulman J, Levine S, Moritz P, Jordan MI, Abbeel P (2015) Trust region policy optimisation. *Proceedings of the 32nd International Conference on Machine Learning* 37: 1889-1897
- Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O (2017) Proximal policy optimisation algorithms arXiv: abs/1707.06347. <https://doi.org/10.48550/arXiv.1707.06347>
- Syromyatnikov VF (1983) *Fundamentals of automation and complex automation of ship's power plants*. (in Rus.). Transport, Moscow
- Theotokatos G (2010) On the cycle mean value modelling of a large two-stroke marine diesel engine. *Proceedings of the Institution of Mechanical Engineers, Part M: Journal of Engineering for the Maritime Environment* 224(3): 193-205. <https://doi.org/10.1243/14750902JEME188>
- Tsujimoto M, Sogihara N, Kuroda M, Sakurada A (2015) Development of a ship performance simulator in actual seas. ASME 2015 34th International Conference on Ocean, Offshore and Arctic Engineering. Newfoundland: ASME OMAE2015-41708, Vo11T12A047. <https://doi.org/10.1115/omae2015-41708>
- van Veldhuizen V (2020) Autotuning PID control using Actor-Critic Deep Reinforcement Learning. Amsterdam: University of Amsterdam. <https://doi.org/10.48550/arXiv.2212.00013>
- Wang H, Ricardez-Sandoval LA (2024) A deep reinforcement learning-based PID tuning strategy for nonlinear MIMO systems with time-varying uncertainty. *IFAC-PapersOnline* 58(14): 887-892. <https://doi.org/10.1016/j.ifacol.2024.08.449>

- Xu XG, Chen XH, Zheng SG (2007) Research on fuzzy PID control to electronic speed regulator. Proceedings of the SPIE 6794: article id. 679407
- Xiros N (2002) Robust control of diesel ship propulsion. Berlin: Springer
- Xiros N (2004) PID marine engine speed regulation under full load conditions for sensitivity H_∞ -norm specifications against propeller disturbance. Journal of Marine Engineering & Technology 3(2): 3-11. <https://doi.org/10.1080/20464177.2004.11020179>
- Zhou Y, Zhang QJ, Zhou QX, Jiang HK (2011) Research on nonlinear PID control of digital governor for diesel engine. In: Zhang W (eds) Advanced Technology in Teaching. Advances in Intelligent and Soft Computing. Berlin: Springer 163: 295-303. https://doi.org/10.1007/978-3-642-29458-7_45