aerospace

MDPI

Article

A Residual PPO Method for Shipboard Helicopter Landing Control

Xiao Chang and Jianliang Ai *

Department of Aeronautics and Astronautics, Fudan University, Shanghai 200433, China; 24210290019@m.fudan.edu.cn
* Correspondence: aijl@fudan.edu.cn

Abstract

Shipboard helicopter landing in the near-deck region requires stable attitude regulation and high-precision deck-relative motion control under substantial model uncertainty and environmental disturbances, conditions under which conventional model-based controllers may lose performance or become overly conservative. This paper proposes a task-oriented, learning-enhanced control algorithm for ship-relative near-deck station keeping and landing by integrating a model-based baseline controller with residual reinforcement learning in a deck-relative closed-loop framework. The algorithmic contribution is the deck-relative baseline–residual control architecture: a split-channel incremental nonlinear dynamic inversion (INDI) outer loop and a reduced-order dynamic inversion (DI) inner loop provide the nominal baseline pathway, while a bounded residual Proximal Policy Optimization (PPO) policy supplies compensation in the same physical outer-loop command channels to suppress unmodeled nonlinearities and time-varying disturbances. Simulation results show that Residual PPO improves hover robustness and landing performance relative to the baseline controller and Pure PPO. With approximately 20–30% residual authority, it achieved 90.0% Desired landing rates in both tested descent-and-landing scenes.

Keywords: shipboard helicopter landing control; ship-relative control; ship airwake; ground effect; residual reinforcement learning; PPO

1. Introduction

Shipboard helicopters play a critical role in maritime reconnaissance, search-and-rescue, and transport missions. Their sortie efficiency and operational safety depend strongly on reliable near-deck landing capability under severe sea conditions [1,2]. In a typical shipboard recovery sequence, the helicopter must complete deck-relative localization, motion-window capture, descent, and touchdown near the flight deck [1,3]. Ship-relative landing control is therefore more than a fixed-point tracking problem. Within the confined near-deck region, the controller must maintain small position and velocity errors, low angular rates, and sufficient control authority to capture a safe landing window [3,4]. Existing ship–helicopter dynamic interface studies further indicate that recovery generally consists of consecutive phases, including stabilized deck-relative holding, lateral translation and alignment, quiescent-period capture, and descent and touchdown [1,4].

Shipboard near-deck landing control is challenging for three closely coupled reasons:

1. The ship–helicopter dynamic interface is intrinsically complex. Wind- and wave-induced ship motion introduces pronounced deck translations and rotations, requiring the helicopter to preserve stability and alignment in a continuously varying deck-relative frame [1,4].

Check for updates

Academic Editor: Bin Zhou

Received: 23 April 2026
Revised: 20 May 2026
Accepted: 29 May 2026
Published: 31 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

2. The near-deck aerodynamic environment is severe. Ship airwake generated by the superstructure, deck geometry, and inflow separation/reattachment, together with near-ground or near-deck effects, produces a strongly unsteady, spatially non-uniform, and direction-dependent flow field above the flight deck, thereby amplifying aerodynamic mismatch and control disturbances [2,5–7].
3. Precision and control requirements are stringent. The confined deck environment demands small position and attitude errors, while high-frequency aerodynamic disturbances, actuator saturation, and limited control authority make accurate and smooth control difficult to achieve simultaneously [2–4].

These coupled difficulties are visualized in Figure 1.

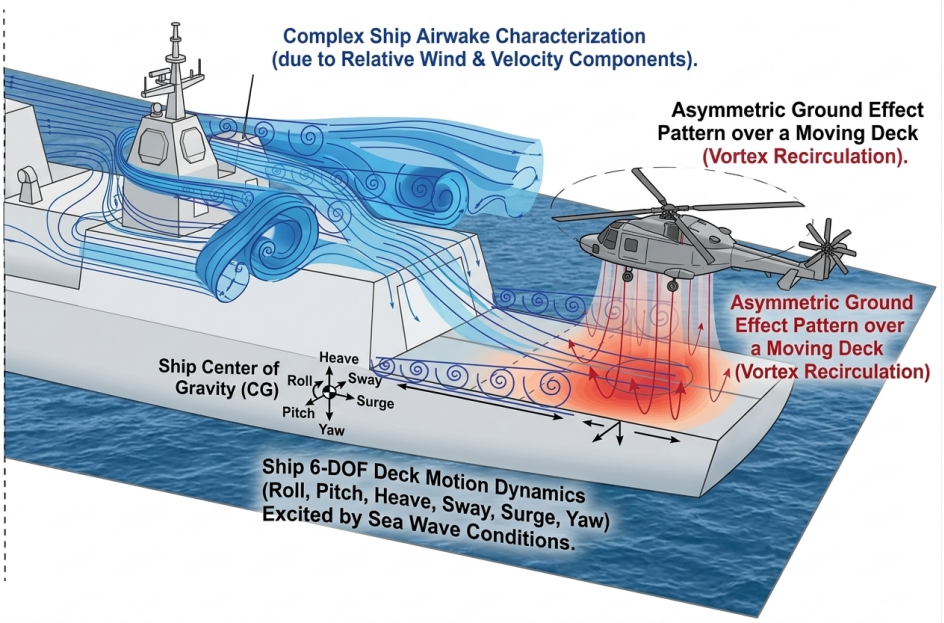

Figure 1. Visualization of the coupled difficulties in near-deck shipboard landing control, including ship motion, airwake interference, and the stringent control requirements imposed on the helicopter during deck-relative operation.

The resulting control problem is accurate and robust deck-relative landing control in a disturbed near-deck environment.

Existing studies on near-deck relative control and shipboard recovery can be broadly grouped into two routes. The first is structured-feedback control based on control-oriented helicopter models. Early work relied on classical optimal designs such as LQR/LQG for automatic approach and ship-deck landing [8,9]. Later studies developed nonlinear and incremental dynamic inversion [10,11], time-delay compensation and disturbance observers [12,13], and fixed-time or state-constrained control [14–16] to improve disturbance rejection and convergence under ship motion and external disturbances. The second route relies on online optimization, most notably model predictive control (MPC). MPC-based methods have been extended from small unmanned helicopters [17] to shipboard operations in ship airwakes and rough seas [18], as well as shrinking- and variable-horizon formulations for touchdown and multi-phase missions [19,20]. These model-based methods provide clear engineering structure, but their effectiveness still depends strongly on model fidelity, disturbance characterization, and online computational capability in strongly unsteady near-deck environments.

Learning-enhanced control offers a complementary route for reducing the model dependence and potential conservatism of structured-feedback and predictive controllers under strongly unsteady near-deck disturbances. A particularly relevant formulation is

the baseline–residual architecture. Johannink et al. formalized residual reinforcement learning as a decomposition in which a conventional controller handles the nominally modelable portion of a task, while a learned policy compensates for the remaining residual component [21]. This decomposition reduces the policy search space without discarding baseline stabilization. Related ideas have also appeared in aerial vehicle control. Liu et al. combined a PID baseline with deep residual reinforcement learning for autonomous blimp control and demonstrated improved robustness to buoyancy variation, wind disturbance, and delay [22]. In another flight-control setting, Yang et al. designed a Proximal Policy Optimization (PPO)-based attitude controller for a tilt-rotor UAV transition process and showed that policy-gradient learning can adapt to pronounced nonlinear aerodynamic variations [23]. These studies support the broader direction of combining model-based control with learning. For ship-relative near-deck landing, however, the learned component should remain confined to residual disturbance compensation rather than replacing the stabilizing baseline.

Motivated by these observations, this paper develops a residual-learning-enhanced control algorithm that combines a split-channel incremental nonlinear dynamic inversion (INDI) outer loop, a reduced-order dynamic inversion (DI) inner loop, and a PPO residual compensator [24]. The novelty of the study is centered on the deck-relative baseline–residual control architecture rather than on redefining the standard INDI, DI, PPO, or residual-learning formulations. The learned residual acts in the same interpretable outer-loop command channels as the INDI/DI baseline, is bounded relative to the nominal control authority, and is trained and evaluated using moving-deck-relative states, rewards, and terminal metrics.

The main contributions of this study are summarized as follows:

1. A deck-relative station-keeping and landing formulation is constructed by expressing the tracking reference, state variables, reward terms, and terminal metrics relative to the moving flight deck.
2. A baseline-preserving residual control algorithm is developed by placing a bounded PPO residual policy on top of a split-channel INDI outer loop and a reduced-order DI inner loop, so that learning augments rather than replaces the nominal control pathway.
3. The integrated controller is evaluated in hover-and-wait and descent-and-landing strong-disturbance conditions. Comparative simulations against the baseline controller, Pure PPO, and additional structured baselines where applicable quantify the benefit and limitations of the residual component under the considered near-deck landing-control scenarios.

2. Problem Formulation

2.1. Helicopter Flight Dynamics Model

The helicopter model used in this study is based on the publicly available J-GenHel nonlinear rotorcraft simulation, which follows the UH-60A/GenHel modeling lineage [25,26]. J-GenHel is a high-fidelity utility-helicopter model that includes a 6 DoF fuselage rigid-body model, nonlinear aerodynamic lookup tables for the fuselage and empennage, rigid main-rotor flap and lead–lag dynamics, a three-state main-rotor inflow model, and a tail-rotor inflow model [25]. The helicopter dynamics are written as

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t), \tag{1}$$

where $\mathbf{x}$ is the state vector and $\mathbf{u}$ is the control input vector. The underlying full-order model contains rigid-body, rotor-flapping, lead–lag, and inflow states, whereas the controller

developed here uses a reduced-order representation that preserves the dominant low-speed near-hover dynamics.

For control-oriented simulation, a 10-state reduced-order model is used in the subsequent controller design, namely

$$\mathbf{x}_{\mathrm{rom}} = \begin{bmatrix} u & v & w & p & q & r & \phi & \theta & \beta_{1c} & \beta_{1s} \end{bmatrix}^{\top}, \tag{2}$$

which retains the dominant rigid-body motion states together with the first-harmonic longitudinal and lateral flapping states. The reduced-order dynamics are written as

$$\dot{\mathbf{x}}_{\mathrm{rom}} = \mathbf{f}_{\mathrm{rom}}(\mathbf{x}_{\mathrm{rom}}, \mathbf{u}, \mathbf{d}, t), \tag{3}$$

where $\mathbf{d}$ denotes the lumped external disturbance and residual modeling uncertainty.

The control input vector is defined as

$$\mathbf{u} = \begin{bmatrix} \delta_{\mathrm{lat}} & \delta_{\mathrm{lon}} & \delta_{\mathrm{col}} & \delta_{\mathrm{ped}} \end{bmatrix}^{\top}, \tag{4}$$

where δ_{lat}, δ_{lon}, δ_{col}, and δ_{ped} denote the lateral cyclic, longitudinal cyclic, collective, and pedal inputs, respectively. Thus, the control-oriented development proceeds from the full nonlinear model in (1) to the reduced state vector (2), the reduced dynamics (3), and the control vector (4). The reduced-order model adopted for controller design was also reported and validated in [25].

The underlying UH-60A/GenHel modeling lineage has long-standing flight-test-based validation in the NASA literature, and the J-GenHel implementation was further reported to reproduce the open-loop dynamics of the corresponding rotorcraft model with good agreement to published flight dynamic characteristics [25–28].

A representative validation result for the adopted flight dynamics model is provided in Figure 2, illustrating the J-GenHel-based helicopter dynamics used in the subsequent control study.

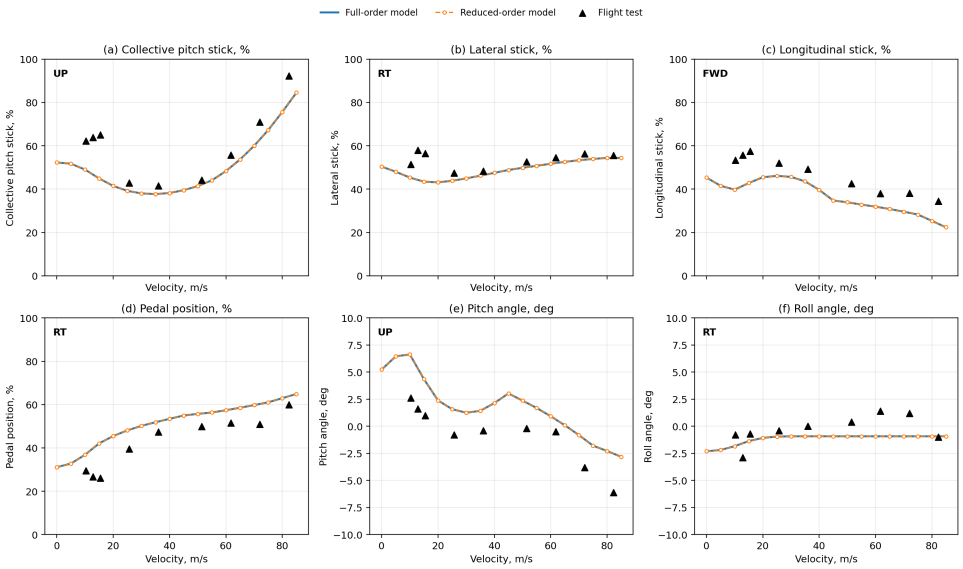

Figure 2. Representative validation image for the adopted helicopter flight dynamics model.

2.2. Control-Oriented Airwake Disturbance Surrogate

To represent near-deck airwake excitation in a compact control-oriented form, this study employs a low-order stochastic shaping-filter surrogate for the longitudinal, lateral, and vertical disturbance components. This modeling choice serves as an engineering

surrogate for near-deck ship-airwake excitation and is consistent with the broader literature on control-equivalent turbulence input modeling for rotorcraft, CETI-based disturbance surrogates, and recent shipboard-operation studies that use compact inflow surrogates for control evaluation [29–33].

The disturbance model is described by the following transfer functions:

$$H_u(s) = \sigma_u \sqrt{\frac{2L_u}{\pi V}} \frac{1}{1 + \frac{L_u}{V}s}, \tag{5}$$

$$H_v(s) = \sigma_v \sqrt{\frac{L_v}{\pi V}} \frac{1 + \sqrt{3}\frac{L_v}{V}s}{\left(1 + \frac{L_v}{V}s\right)^2}, \tag{6}$$

$$H_w(s) = \sigma_w \sqrt{\frac{L_w}{\pi V}} \frac{1 + \sqrt{3}\frac{L_w}{V}s}{\left(1 + \frac{L_w}{V}s\right)^2}, \tag{7}$$

where $H_u(s)$, $H_v(s)$, and $H_w(s)$ denote the shaping filters for the longitudinal, lateral, and vertical airwake components, respectively; V is the reference flight speed; L_u, L_v, and L_w are the turbulence scale lengths; and σ_u, σ_v, and σ_w are the corresponding turbulence intensities.

Driven by mutually independent unit white-noise inputs, the CETI disturbance vector is written as

$$\mathbf{d}_{\mathrm{aw}}(t) = \begin{bmatrix} u_{\mathrm{aw}}(t) & v_{\mathrm{aw}}(t) & w_{\mathrm{aw}}(t) \end{bmatrix}^{\top}, \tag{8}$$

with

$$u_{\mathrm{aw}}(s) = H_u(s)n_u(s), \qquad v_{\mathrm{aw}}(s) = H_v(s)n_v(s), \qquad w_{\mathrm{aw}}(s) = H_w(s)n_w(s), \tag{9}$$

where $n_u(s)$, $n_v(s)$, and $n_w(s)$ are the white-noise excitation signals.

In the present work, the shaping filters (5)–(7) generate the disturbance vector (8) through the component definitions in (9). The resulting disturbance vector $\mathbf{d}_{\mathrm{aw}}$ is introduced into the near-deck helicopter model as the external airwake input. Representative time histories generated by the CETI model are shown in Figure 3.

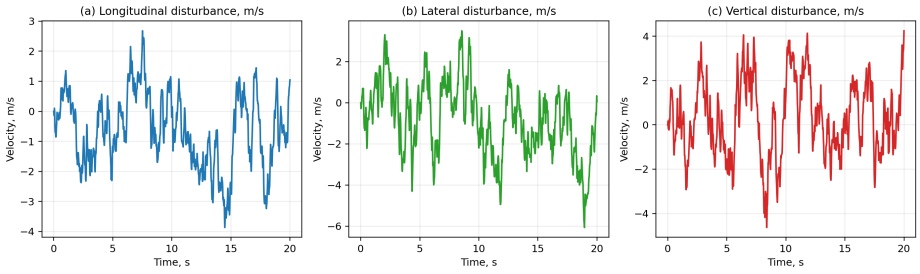

Figure 3. Representative time histories of the CETI-generated ship-airwake disturbance components in the longitudinal, lateral, and vertical directions under white-noise excitation.

2.3. Ground-Effect Model

To represent near-deck in-ground-effect (IGE) behavior during shipboard landing control, this study adopts the classical Cheeseman–Bennett ground-effect model [34]. For a

rotor operating at constant power in a quasi-steady low-speed condition, the thrust ratio in and out of ground effect is written as

$$\left.\frac{T_{\mathrm{IGE}}}{T_{\mathrm{OGE}}}\right|_{P=\mathrm{const}} = \frac{1}{1-\left(\frac{R_m}{4z_h}\right)^2} = \frac{1}{k_G(z_h)}, \tag{10}$$

where T_{IGE} and T_{OGE} denote the rotor thrust in and out of ground effect, respectively, R_m is the main-rotor radius, and z_h is the instantaneous rotor-hub height above the deck. Because the present study focuses on low-speed near-deck ship-relative landing control, Equation (10) is adopted as the engineering correction used to modify the rotor thrust according to the instantaneous hub-to-deck clearance z_h.

This ground-effect representation is simplified. The quasi-steady correction model adopted here cannot accurately capture dynamic wake-inflow effects, inclined-ground effects, or moving-deck-induced inflow dynamics, and therefore cannot fully reproduce rotor ground-effect aerodynamics under complex operating conditions. Recent studies by Pasquali et al. have provided systematic numerical and experimental investigations of hovering-rotor ground effect over fixed and moving ground, including inclined-surface conditions and state-space dynamic inflow modelling [35,36]. These studies provide a basis for future improvements to the present model. The simplified correction is retained here to reduce simulation complexity and focus on controller-algorithm verification. Future research can incorporate higher-fidelity ground-effect and flight dynamics models that include dynamic wake, inclined ground, and moving-deck dynamic inflow effects to improve model accuracy and operating-envelope coverage.

2.4. Ship-Deck Motion Model

A ship undergoes six-degree-of-freedom motion under wave excitation, including surge, sway, heave, roll, pitch, and yaw. For the present near-deck landing-control task, the deck-motion components that most directly affect the relative landing condition are heave, pitch, and roll. A simplified sinusoidal superposition model is therefore adopted to simulate the deck motion of a representative ship [37].

The ship heave motion is modeled as

$$h_s(t) = 0.2172\sin(0.4t) + 0.4714\sin(0.5t) + 0.3592\sin(0.6t) + 0.2227\sin(0.7t), \tag{11}$$

the ship pitch motion is modeled as

$$\theta_s(t) = 0.005\sin(0.46t) + 0.00946\sin(0.58t) + 0.00725\sin(0.7t) + 0.00845\sin(0.82t), \tag{12}$$

and the ship roll motion is modeled as

$$\phi_s(t) = 0.021\sin(0.46t) + 0.0431\sin(0.54t) + 0.029\sin(0.62t) + 0.022\sin(0.67t). \tag{13}$$

The corresponding vertical and lateral motions of the landing point are approximated as

$$h_p(t) = 57.11\sin\theta_s(t) + 13.2\sin^2(0.5\phi_s(t)) + h_s(t), \tag{14}$$

and

$$y_p(t) = 6.6\sin\phi_s(t), \tag{15}$$

where $h_s(t)$, $\theta_s(t)$, and $\phi_s(t)$ denote the heave, pitch, and roll motions of the ship, respectively; $h_p(t)$ and $y_p(t)$ denote the vertical and lateral motions of the landing point on the deck; and t is time.

Equations (11)–(15) are used to generate the time-varying deck-motion disturbance and the associated landing-point motion introduced into the relative-motion model and control framework. The corresponding analytic multisine ship motion histories are shown in Figure 4, which visualizes the deck-motion model adopted in this study.

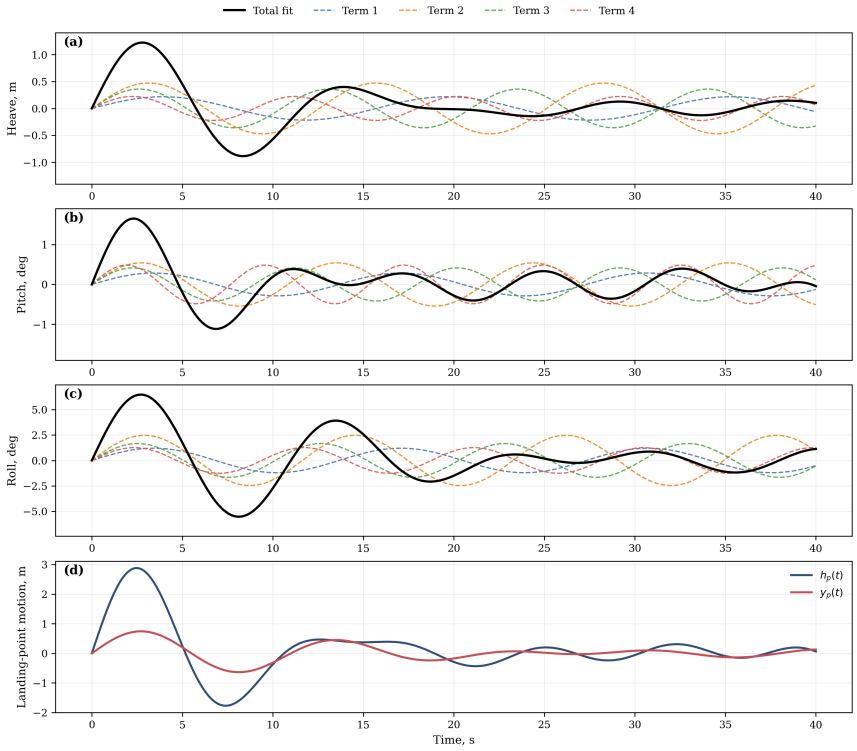

Figure 4. Analytic multisine ship motion model used in this study, showing the generated heave (**a**), pitch (**b**), and roll (**c**) motions together with the resulting landing-point vertical and lateral motions (**d**).

3. Methods

3.1. Overall Architecture

The integrated controller for ship-relative near-deck landing control uses a baseline–residual outer-loop architecture together with a reduced-order dynamic inversion inner loop. The split-channel INDI baseline provides the nominal stabilization pathway and generates an interpretable outer-loop command from deck-relative tracking errors. The PPO policy outputs only an additive residual command for disturbance compensation, so the nominal stabilization layer is preserved. The combined outer-loop command is then passed to the dynamic inversion inner loop for rapid realization [10,11,21,24].

The INDI/DI and PPO blocks follow established formulations, but their coupling here defines a task-specific control algorithm for moving-deck operation. The key design choices are the deck-relative observation and error variables, the four-channel residual action interface, the residual scaling relative to the nominal outer-loop authority, and the strong-disturbance evaluation scenarios used to test station keeping and landing.

Within this closed-loop framework, the deck-motion input is generated by the analytic multisine model introduced in Section 2.4 and visualized in Figure 4.

In this structure, d denotes the CETI disturbance input, $\boldsymbol{u}_b$ is the baseline outer-loop command, $\Delta\boldsymbol{u}_{\mathrm{rl}}$ is the residual command generated by the PPO policy, and $\boldsymbol{u}_{\mathrm{cmd}}$ is the final composite outer-loop command delivered to the inner-loop controller. The "task command" denotes the deck-fixed relative-position reference, whereas "near-deck helicopter dynamics" denotes the nonlinear plant to be controlled.

The baseline controller maintains nominal deck-relative control performance, the PPO policy learns only the residual component associated with unmodeled nonlinearities and time-varying disturbances, and the inner loop converts the composite outer-loop command into realizable plant inputs. The resulting signal flow is summarized in Figure 5.

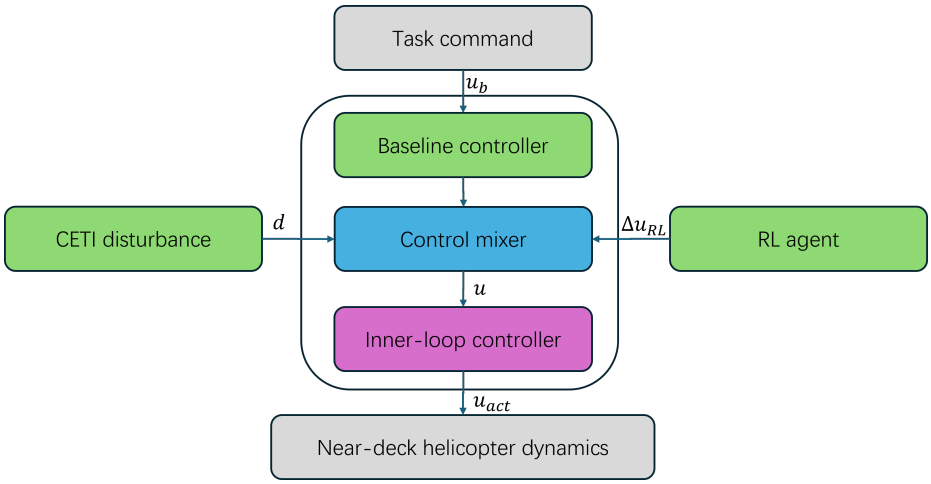

Figure 5. Overall control-framework flow of the baseline–residual architecture used for the ship-relative near-deck landing-control task.

3.2. Baseline Outer-Loop Law

The baseline controller used in the reported experiments is a split-channel INDI outer loop that serves as the nominal deck-relative controller in both the baseline-controller configuration and the Residual PPO architecture [10,11]. In the active implementation, the effective outer-loop output is not a direct actuator command but the raw outer-command vector

$$\boldsymbol{u}_b = \begin{bmatrix} \phi_c & \theta_c & v_{z,c} & r_c \end{bmatrix}^{\mathsf{T}}, \tag{16}$$

where ϕ_c and θ_c are commanded roll and pitch, $v_{z,c}$ is the commanded inertial vertical speed, and r_c is the commanded yaw rate.

The baseline law is built from deck-relative position, velocity, and heading errors with integral action. Only the channel laws most relevant to the reported controller are retained. The horizontal outer loop is intentionally asymmetric. In the longitudinal channel, the desired x-axis pseudo-acceleration is converted into a pitch command through a θ-based INDI law,

$$\theta_c = \theta_{\text{trim}} + \theta_{\text{bias}} + k_{\text{INDI},x} \frac{a_{\text{des},x} - a_{x,\text{sync}}}{g_{\theta a_x}}, \tag{17}$$

where $a_{\text{des},x}$ is the desired longitudinal pseudo-acceleration obtained from deck-relative tracking errors and $a_{x,\text{sync}}$ is the synchronized measured longitudinal acceleration. By contrast, the lateral channel uses a guidance-type roll command rather than direct acceleration inversion:

$$\phi_c = \phi_{\text{trim}} + \phi_{\text{bias}} + \text{sat}\left(k_{p,y} e_y + k_{d,y} e_{vy} + k_{i,y} \eta_y, \bar{\phi}_{\text{guid}}\right), \tag{18}$$

In the vertical channel, the outer loop generates an inertial vertical-speed command rather than a direct collective input,

$$v_{z,c} = \text{sat}\left(v_{z,\text{ff}} + k_{p,z} e_z + k_{d,z} e_{vz} + k_{i,z} \eta_z - z_{\text{hedge}}, \bar{v}_z\right), \tag{19}$$

while the yaw channel uses a conservative yaw-rate hold law,

$$r_c = k_\psi e_\psi - d_\psi r. \tag{20}$$

The active baseline is therefore defined by four structured outer-loop channels: longitudinal θ-based INDI, lateral roll guidance, vertical inertial-speed command generation, and yaw-rate hold. In the reported experiments, the raw split-channel output (16) is used directly as the nominal command for the residual formulation described next.

3.3. MDP Formulation and PPO Residual Learning

The learned policy is designed to modify only the structured outer-loop command produced by the nominal baseline, rather than replacing the baseline controller itself. This choice follows the residual-reinforcement-learning principle that the model-based controller handles the nominally modelable part of the waiting-task control problem, while the learned policy compensates for the remaining disturbance-driven mismatch and modeling residuals [21–23].

Let the raw policy output at time step k be denoted by

$$\tilde{\boldsymbol{a}}_k \in \mathbb{R}^4. \tag{21}$$

To keep the learned correction bounded in the same four outer-loop channels as the baseline command, the policy output is clipped component-wise to

$$\boldsymbol{a}_{\mathrm{rl},k} = \mathrm{sat}\Big(\tilde{\boldsymbol{a}}_k, [-1,1]^4\Big), \tag{22}$$

and then mapped into the physical command space through a fixed diagonal scaling matrix,

$$\Delta \boldsymbol{u}_{\mathrm{rl},k} = \boldsymbol{S}_r \boldsymbol{a}_{\mathrm{rl},k}, \tag{23}$$

where

$$\boldsymbol{S}_r = \mathrm{diag}(s_1, s_2, s_3, s_4). \tag{24}$$

The final command sent to the inner loop is therefore

$$\boldsymbol{u}_{\mathrm{cmd},k} = \boldsymbol{u}_{b,k} + \Delta \boldsymbol{u}_{\mathrm{rl},k}. \tag{25}$$

Equations (22)–(25) define the Residual PPO structure. The baseline command $\boldsymbol{u}_b$ preserves the nominal deck-relative tracking semantics, whereas the learned policy is restricted to bounded additive correction in the roll, pitch, inertial vertical-speed, and yaw-rate channels.

For comparison, a Pure PPO controller is also evaluated. It uses the same reduced-order DI inner loop and the same four outer-loop command channels, but removes the additive baseline pathway. In the released comparison artifact, this policy directly generates

$$\boldsymbol{u}^{\mathrm{ppo}}_{\mathrm{cmd},k} = \mathrm{diag}(0.28,\ 0.28,\ 0.24,\ 1.00)\ \mathrm{sat}\Big(\tilde{\boldsymbol{a}}^{\mathrm{ppo}}_k, [-1,1]^4\Big), \tag{26}$$

where the four channels again correspond to commanded roll, pitch, inertial vertical speed, and yaw rate. The observation retains the core deck-relative tracking variables and previous-command history, but omits baseline-specific quantities such as $\boldsymbol{u}_b$, the residual history, and mixer variables. This comparison isolates the effect of preserving the nominal model-based command pathway.

The residual-learning problem is formulated as a finite-horizon Markov decision process (MDP)

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma), \tag{27}$$

where $\mathcal{S}$ and $\mathcal{A}$ denote the observation and action spaces, $\mathcal{P}$ is the transition kernel induced by the helicopter dynamics together with the baseline generator and DI inner loop, r is the scalar reward, and $\gamma \in (0, 1)$ is the discount factor.

The observation vector contains the variables most directly related to deck-relative correction and residual-command generation:

$$\boldsymbol{s}_k = \begin{bmatrix} \boldsymbol{e}_{p,k}^{\mathsf{T}} & \boldsymbol{e}_{v,k}^{\mathsf{T}} & \boldsymbol{e}_{\mathrm{att},k}^{\mathsf{T}} & \boldsymbol{\eta}_k^{\mathsf{T}} & p_k & q_k & r_k & \boldsymbol{u}_{b,k}^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}}, \tag{28}$$

where $\boldsymbol{e}_{p,k}$, $\boldsymbol{e}_{v,k}$, and $\boldsymbol{e}_{\mathrm{att},k}$ denote the deck-relative position, velocity, and attitude tracking errors, respectively; $\boldsymbol{\eta}_k$ is the outer-loop integral state; p_k, q_k, and r_k are the body-axis angular rates; and $\boldsymbol{u}_{b,k}$ is the nominal baseline outer-loop command. Including $\boldsymbol{u}_{b,k}$ in the observation makes the residual policy explicitly conditioned on the baseline intent, so that the learned action is interpreted as context-dependent correction around the nominal command rather than as an independent end-to-end control law.

The action space is defined by the bounded normalized residual command

$$\boldsymbol{a}_k = \boldsymbol{a}_{\mathrm{rl},k} \in [-1, 1]^4, \tag{29}$$

and PPO is used to optimize a stochastic policy $\pi_\theta(\boldsymbol{a}_k|\boldsymbol{s}_k)$ together with a critic $V_\phi(\boldsymbol{s}_k)$ that estimates the state value. The discounted-return objective is

$$J(\pi_\theta) = \mathbb{E}_{\pi_\theta}\left[\sum_{k=0}^{T-1} \gamma^k r_k\right]. \tag{30}$$

For compactness, the reward is introduced here in schematic form as

$$r_k = -w_p\|\boldsymbol{e}_{p,k}\|_2^2 - w_v\|\boldsymbol{e}_{v,k}\|_2^2 - w_a\|\boldsymbol{e}_{\mathrm{att},k}\|_2^2 - w_u\|\Delta\boldsymbol{u}_{\mathrm{rl},k}\|_2^2 + r_k^{\mathrm{term}}, \tag{31}$$

where the first three terms penalize deck-relative tracking error, the fourth term penalizes excessive residual-command usage, and r_k^{term} collects terminal bonuses and penalties. Equation (31) summarizes the reward-design intent. The implemented reward components and numerical coefficients used in the reported training runs are given in Section 4.1.

Following PPO, policy updates are performed on rollout batches collected under the current closed-loop controller [24]. Let $\pi_{\theta_{\mathrm{old}}}$ denote the behavior policy used to collect the rollout data. The importance ratio is written as

$$\rho_k(\theta) = \frac{\pi_\theta(\boldsymbol{a}_k|\boldsymbol{s}_k)}{\pi_{\theta_{\mathrm{old}}}(\boldsymbol{a}_k|\boldsymbol{s}_k)}. \tag{32}$$

Using the estimated advantage $\hat{A}_k$, the clipped PPO surrogate objective becomes

$$L_\pi(\theta) = \mathbb{E}\left[\min\left(\rho_k(\theta)\hat{A}_k,\ \mathrm{clip}\left(\rho_k(\theta), 1-\epsilon, 1+\epsilon\right)\hat{A}_k\right)\right], \tag{33}$$

where $\epsilon > 0$ is the clip parameter. To encourage sufficient exploration during training, an entropy bonus is added in the practical implementation,

$$L_\pi^{\mathrm{ent}}(\theta) = L_\pi(\theta) + c_e\,\mathbb{E}[H(\pi_\theta(\cdot|\boldsymbol{s}_k))], \tag{34}$$

where $H(\cdot)$ denotes policy entropy and c_e is the entropy coefficient.

The value function is trained to approximate the expected discounted return from the current state,

$$V_\phi(\boldsymbol{s}_k) \approx \mathbb{E}_{\pi_\theta}\left[\sum_{l=0}^{T-k-1} \gamma^l r_{k+l} \,\middle|\, \boldsymbol{s}_k\right]. \tag{35}$$

Advantage estimation is computed by generalized advantage estimation (GAE) [38],

$$\hat{A}_k = \sum_{l=0}^{T-k-1} (\gamma\lambda)^l \, \delta_{k+l}, \tag{36}$$

with temporal-difference residual

$$\delta_k = r_k + \gamma V_\phi(\boldsymbol{s}_{k+1}) - V_\phi(\boldsymbol{s}_k), \tag{37}$$

where $\lambda \in [0, 1]$ is the trace-decay parameter. The corresponding value target is

$$\hat{R}_k = \hat{A}_k + V_\phi(\boldsymbol{s}_k), \tag{38}$$

and the critic is trained by squared-error regression,

$$L_V(\phi) = \mathbb{E}\Big[\big(V_\phi(\boldsymbol{s}_k) - \hat{R}_k\big)^2\Big]. \tag{39}$$

The learning procedure can therefore be interpreted as follows. The baseline controller first generates a structured nominal holding command in physically meaningful outer-loop channels. The PPO actor then proposes only a bounded residual correction in the same channels, while the critic evaluates the long-horizon benefit of that correction under the closed-loop near-deck relative-motion dynamics. This decomposition reduces the burden on the learned policy relative to end-to-end control and is consistent with the baseline-preserving residual-learning philosophy adopted here.

3.4. DI Inner-Loop Realization

The composite outer-loop command in (25) is passed to a reduced-order dynamic inversion inner loop that realizes the commanded heave velocity and body-axis angular-rate response. The key vertical-channel transformation is the conversion of the inertial vertical-speed command into a body-axis heave reference:

$$w_{\text{ref}} = \text{sat}\left(\frac{v_{z,c} + u\sin\theta - v\sin\phi\cos\theta}{\cos\phi\cos\theta}, \bar{w}_{\text{ref}}\right), \tag{40}$$

where denominator protection is used in the implementation and $\bar{w}_{\text{ref}}$ denotes the heave-reference limit. Thus, the third outer-loop component is interpreted as an inertial vertical-speed target and then tracked by the DI layer as a body-axis heave-velocity reference.

With fast z-axis DI enabled, the reduced-order inner loop closes the four variables $[w, p, q, r]^\mathsf{T}$ through a compact inversion written as

$$\Delta\boldsymbol{u} = \boldsymbol{C}_{B,wpqr}^{-1}\left(\boldsymbol{\nu}_{wpqr} - \boldsymbol{C}_{A,wpqr}\begin{bmatrix} w & p & q & r \end{bmatrix}^\mathsf{T}\right), \tag{41}$$

where $\boldsymbol{\nu}_{wpqr}$ denotes the pseudo-control vector for the heave-velocity and body-rate channels, and $\Delta\boldsymbol{u} = [\Delta u_{\text{lat}}, \Delta u_{\text{lon}}, \Delta u_{\text{col}}, \Delta u_{\text{ped}}]^\mathsf{T}$ is the corresponding control increment. The final plant input is then

$$\boldsymbol{u}_{\text{plant}} = \begin{bmatrix} \Delta u_{\text{lat}} + \delta_{\text{lat},f} \\ \Delta u_{\text{lon}} + \delta_{\text{lon},f} \\ \Delta u_{\text{col}} \\ \Delta u_{\text{ped}} + \delta_{\text{ped},f} \end{bmatrix}, \tag{42}$$

where $\delta_{\text{lat},f}$, $\delta_{\text{lon},f}$, and $\delta_{\text{ped},f}$ denote the actuator-like internal states of the DI implementation. This compact mapping is sufficient for the present study because the learned policy

corrects the outer-loop command, whereas the DI layer remains responsible for generating realizable plant inputs.

The resulting controller is therefore characterized as a raw split-channel INDI outer loop feeding a reduced-order WPQR dynamic inversion inner loop with fast heave-axis DI.

4. Results

4.1. Simulation Setup and Metrics

To provide a unified and transparent performance basis for the hover-and-wait strong-disturbance condition, the results are interpreted with reference to the ADS-33E-PRF mission-task-element (MTE) criteria for the Cargo/Utility class [39]. Because the present platform follows the UH-60A utility-helicopter modeling lineage, the Desired and Adequate performance envelopes are adopted as benchmark criteria for sustained ship-relative hover holding. ADS-33E-PRF is used here as a structured low-speed performance reference rather than as a complete shipboard airworthiness certification standard.

In addition to ADS-33 Desired/Adequate occupancy, a controller-level success-hold metric is reported. A run is counted as success-hold compliant if, with the quiescent gate active, the controller success tube $\|\boldsymbol{e}_p\|_2 \le 0.60$ m, $\|\boldsymbol{e}_v\|_2 \le 0.35$ m/s, and $\|\boldsymbol{e}_{\mathrm{att}}\|_2 \le 5^\circ$ is maintained continuously over the full evaluation segment. For the multi-WOD comparison, the mean command-usage metric is defined as the episode-mean normalized command magnitude,

$$U_{\mathrm{usage}} = 100\,\frac{1}{N}\sum_{k=1}^{N}\frac{1}{4}\sum_{i=1}^{4}\mathrm{clip}\!\left(\frac{|u_{\mathrm{cmd},i}(k)|}{u_{i,\max}},0,\,1\right), \tag{43}$$

where $u_{i,\max}$ denotes the controller-specific channel-wise command limit and N is the number of decision steps in the episode. For Baseline and Residual PPO, these limits are the outer-loop command bounds; for Pure PPO, they are the direct action scales of the four commanded channels.

The PPO training details used for the reported residual controller are summarized in Table 1. The reward used in the reported residual-controller training is

$$r_t = r_{\mathrm{err}} + r_{\mathrm{res}} + r_{\mathrm{smooth}} + r_{\mathrm{progress}} + r_{\mathrm{near}} + r_{\mathrm{hold}} + r_{\mathrm{terminal}}, \tag{44}$$

where the principal shaped components are given by

$$r_{\mathrm{err}} = -1.50\|\boldsymbol{e}_p\|_2^2 - 0.80\|\boldsymbol{e}_v\|_2^2 - 0.20\|\boldsymbol{e}_{\mathrm{att}}\|_2^2 - 0.05\|\boldsymbol{\omega}\|_2^2, \tag{45}$$

$$r_{\mathrm{res}} = -0.0025\|\Delta\boldsymbol{u}_{\mathrm{rl},t}\|_2^2, \tag{46}$$

$$r_{\mathrm{smooth}} = -0.02\|\boldsymbol{u}_{\mathrm{cmd},t} - \boldsymbol{u}_{\mathrm{cmd},t-1}\|_2^2, \tag{47}$$

$$r_{\mathrm{progress}} = 0.03\,\mathrm{clip}\!\left(\frac{e_{\mathrm{outer},t-1} - e_{\mathrm{outer},t}}{\max(e_{\mathrm{outer},0},\,0.25)},-1,\,1\right), \tag{48}$$

$$r_{\mathrm{near}} = \begin{cases} 0.04, & \|\boldsymbol{e}_p\|_2 \le 0.12,\ \|\boldsymbol{e}_v\|_2 \le 0.35, \\ 0.02, & \|\boldsymbol{e}_p\|_2 \le 0.25,\ \|\boldsymbol{e}_v\|_2 \le 1.5\times 0.35, \\ 0.01, & \|\boldsymbol{e}_p\|_2 \le 0.50,\ \|\boldsymbol{e}_v\|_2 \le 2.0\times 0.35, \\ 0, & \text{otherwise}, \end{cases} \tag{49}$$

$$r_{\mathrm{hold}} = \mathbb{I}_{\mathrm{tube}}\left(0.06 + 0.18\,\mathrm{clip}\!\left(\frac{n_{\mathrm{hold}}}{250},0,1\right)\right), \tag{50}$$

$$r_{\mathrm{terminal}} = 25\,\mathbb{I}_{\mathrm{succ}} - 18\,\mathbb{I}_{\mathrm{fail}}, \tag{51}$$

and

$$e_{\text{outer},t} = \|\boldsymbol{e}_{p,t}\|_2. \tag{52}$$

Here, $\boldsymbol{e}_p$ is expressed in meters, $\boldsymbol{e}_v$ in meters per second, $\boldsymbol{e}_{\text{att}}$ in radians, and $\boldsymbol{\omega}$ in radians per second. The indicator $\mathbb{I}_{\text{tube}}$ is one when the success tube is active, $\mathbb{I}_{\text{succ}}$ is one at successful episode termination, $\mathbb{I}_{\text{fail}}$ is one at failure termination, and n_{hold} denotes the consecutive in-tube decision-step count.

Table 1. PPO training details for the reported residual controller.

Parameter	Value
Actor/critic learning rate	1.5×10^{-4}
Total training timesteps	2.6×10^{5}
Rollout length (n_{steps})	2048
PPO epochs per update	10
Discount factor γ	0.995
GAE lambda λ_{GAE}	0.97
PPO clip range	0.2
Entropy coefficient	5×10^{-4}
Value-loss coefficient	0.5
Maximum gradient norm	0.5
Target KL	0.02
Actor network (hidden layers/neurons)	2/(256, 128)
Critic network (hidden layers/neurons)	2/(256, 128)
Batch size	256

The fixed residual-command scaling is summarized in Table 2. Because the normalized residual action is bounded in $[-1, 1]^4$ before the fixed scaling matrix is applied, each scale entry is also the corresponding channel-wise residual upper limit. Relative to the nominal outer-loop authority used by the baseline channels, these limits correspond to an effective PPO residual authority of approximately 20–30%. This range provides a practical compromise: smaller authority reduces disturbance-compensation capability, whereas substantially larger authority allows the learned residual to dominate the baseline pathway and can weaken out-of-distribution robustness.

Table 2. Fixed residual-command scaling.

Channel	Command Variable	Unit	Scale/Upper Limit
Roll	ϕ_c	rad	0.12
Pitch	θ_c	rad	0.12
Vertical speed	$v_{z,c}$	m/s	0.08
Yaw rate	r_c	rad/s	0.05

4.2. Representative Deck-Relative Response Histories

To evaluate waiting-stage performance, the helicopter is commanded to maintain a deck-fixed near-deck reference under the representative disturbance condition defined above. Figure 6 shows the deck-relative flight-state responses of the three controllers, including relative height, height error, angle of attack, speed, pitch angle, and pitch rate.

Under this representative disturbance condition, all three controllers maintain near-deck hover, but with clearly different transient behavior. Residual PPO yields smaller oscillations, weaker overshoot, and faster error convergence than both the baseline and Pure PPO controllers, indicating improved compensation of near-deck disturbances in this test case. The angle-of-attack and speed responses show the same trend. Overall, Residual PPO provides the most favorable deck-relative tracking and disturbance-rejection metrics

among the three tested controllers, while the baseline controller remains competitive and Pure PPO still completes the task.

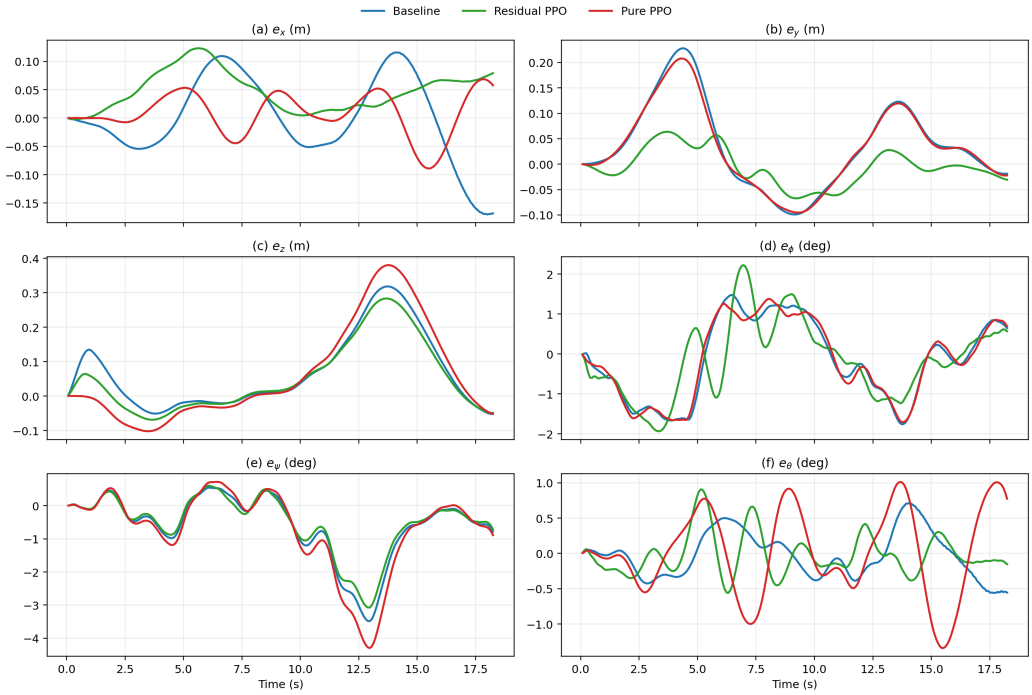

Figure 6. Flight-state response curves of the three controllers in the ship-relative waiting task: (**a**) relative-height response; (**b**) height-error response; (**c**) angle-of-attack response; (**d**) speed response; (**e**) pitch-angle response; and (**f**) pitch-rate response.

A formal classification of sustained hover quality in terms of the ADS-33E-based Desired and Adequate envelopes is provided in the following Monte Carlo evaluation.

4.3. Monte Carlo and WOD Robustness Evaluation

To provide a broader assessment of the three controllers in the ship-relative waiting task, a Monte Carlo simulation study was conducted. For the fixed-WOD comparison, 500 randomized trials were performed for each control scheme, namely Pure PPO, the baseline controller, and Residual PPO, at 0° wind-over-deck (WOD), with randomized CETI disturbance seeds and hover-initialization errors. All trials used the same waiting-task setting, with only the disturbance realization and initial deck-relative error randomized, to evaluate closed-loop robustness under stochastic near-deck disturbances [40].

The corresponding aggregate Monte Carlo counts and rates are summarized in Table 3.

Table 3. Monte Carlo performance statistics under the ADS-33E-based Desired and Adequate criteria.

Metric	Baseline	Pure PPO	Residual PPO
Desired count	356	288	**393**
Adequate count	445	417	**477**
Desired rate (%)	71.2	57.6	**78.6**
Adequate rate (%)	89.0	83.4	**95.4**
Criteria-satisfying hold rate (%)	93.2	86.8	**96.6**
Median terminal error (m)	0.333	0.406	**0.241**
95th-percentile position error (m)	1.235	1.232	**1.227**

The Monte Carlo outcomes are classified according to the ADS-33E-PRF Hover MTE criteria for the Cargo/Utility class [39]. A continuous hold duration of 30 s is used as the evaluation window. Desired performance is defined as maintaining the longitudinal

and lateral position errors within ±3 ft, the altitude error within ±2 ft, and the heading error within ±5°. Adequate performance is defined as maintaining the longitudinal and lateral position errors within ±6 ft, the altitude error within ±4 ft, and the heading error within ±10°. The adequate longitudinal/lateral hover tolerance adopted here is ±6 ft, consistent with the Cargo/Utility column of ADS-33E-PRF rather than the Scout/Attack DVE criterion [39].

Figure 7 presents the terminal statistical distributions of the three controllers under these ADS-33E-based criteria. Among 500 trials, Pure PPO satisfied the Desired and Adequate criteria in 288 and 417 trials, respectively, while the corresponding numbers for the baseline controller were 356 and 445. Residual PPO achieved the highest occupancy of the Desired region, with 393 trials satisfying the Desired criterion and 477 trials satisfying the Adequate criterion. Residual PPO also achieved the highest criteria-satisfying hold rate (96.6%), the lowest median terminal position error (0.241 m), and the lowest 95th-percentile position error (1.227 m). These results indicate that Residual PPO provides the most favorable task-level deck-relative holding metrics among the tested controllers under randomized shipboard disturbances. The baseline controller also gives higher Desired/Adequate occupancy and hold rate than Pure PPO, but remains below Residual PPO on the same evaluation set.

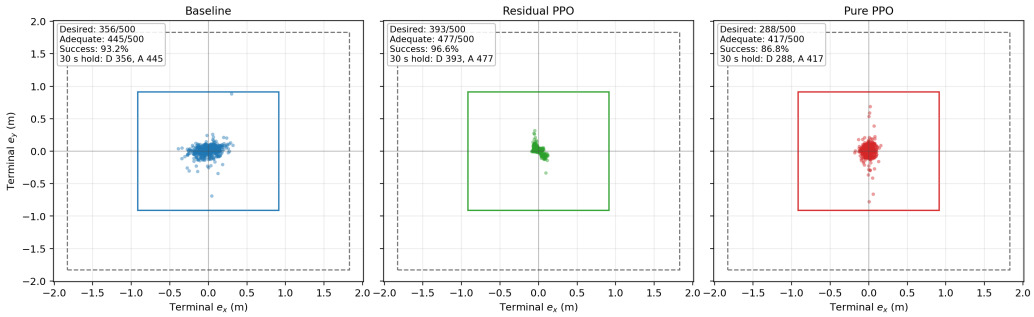

Figure 7. Monte Carlo terminal distributions of the three controllers in the ship-relative waiting task under the ADS-33E-based criteria over 500 randomized fixed-WOD (0°) trials.

To further examine sensitivity to wind direction, an additional wind-over-deck (WOD) sweep was carried out at 0°, 45°, and 90°. Figure 8 reports the Monte Carlo distributions of deck-relative position RMS over 500 randomized runs per controller and per WOD direction, and Table 4 summarizes the corresponding success, XYZ RMS, Desired, Adequate, and command-usage statistics.

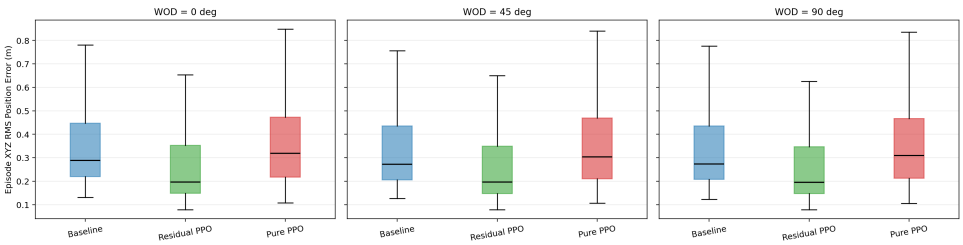

Figure 8. Monte Carlo WOD-sweep comparison for the ship-relative waiting task. The results summarize 500 randomized runs per controller and per wind-over-deck direction.

The WOD-sweep results in Figure 8 and Table 4 show the expected generalization difference between pure and residual learning. At the trained 0° WOD condition, Pure PPO remains comparable to the structured baseline in Desired and Adequate occupancy. Once the wind direction moves outside the training condition, however, Pure PPO degrades sharply, with Desired rates falling to 22.4% and 18.6% at 45° and 90°, respectively. Residual PPO

maintains 84.8–86.0% Desired rates and the lowest RMS tracking error across all three WOD settings, while also using less command activity. These data suggest that residual learning improves practical hover-hold performance by using the PPO component as a bounded compensator rather than as a full replacement for the nominal model-based pathway.

Table 4. Multi-WOD Monte Carlo summary for the ship-relative waiting task.

Controller	WOD (°)	Success (%)	XYZ RMS (m)	Desired (%)	Adequate (%)	Command Usage (%)
Baseline	0	93.0	0.342	70.8	90.2	31.6
Residual PPO	0	**97.0**	**0.246**	**86.0**	**96.6**	**28.5**
Pure PPO	0	92.2	0.346	70.2	89.8	34.6
Baseline	45	92.4	0.351	68.4	89.6	31.8
Residual PPO	45	**96.4**	**0.257**	**84.8**	**95.8**	**28.6**
Pure PPO	45	70.0	0.628	22.4	64.0	42.1
Baseline	90	92.0	0.357	67.8	89.0	32.0
Residual PPO	90	**96.2**	**0.263**	**85.4**	**95.6**	**28.8**
Pure PPO	90	66.6	0.704	18.6	60.2	44.7

4.4. Descent-and-Landing Strong-Disturbance Condition

The preceding results evaluate the hover-and-wait strong-disturbance condition. To examine whether the same learning-enhanced architecture remains useful during descent, an additional descent-and-landing strong-disturbance condition was evaluated. The first scene retains the original disturbance setting, namely the original wind direction, CETI intensity, ship motion, and dynamic moving-deck ground-effect model, but replaces the holding task with a 10 s descent-and-landing task. The second scene is a moving-deck descent scene. The compared controllers are Residual PPO, Pure PPO, Pure INDI, SHMPC, and DOB/CTSMC.

For the descent-and-landing strong-disturbance condition, the hover Desired/Adequate ADS-33E interpretation is not reused as a landing certification standard. Instead, terminal landing classes are defined directly from the final deck-relative position, velocity, and attitude errors. In the original-disturbance descent scene, Desired requires a terminal position error below 0.35 m, a terminal velocity error below 0.30 m/s, and a terminal attitude error below 3.5°. In the moving-deck descent scene, where the terminal alignment requirement is more stringent, Desired requires a terminal position error below 0.185 m, a terminal velocity error below 0.34 m/s, and a terminal attitude error below 1.0°. For both scenes, Adequate requires a terminal position error below 5.0 m, a terminal velocity error below 2.0 m/s, and a terminal attitude error below 15°. A terminal score

$$J_T = \|e_p(T)\| + 4\|e_v(T)\| + 0.2\|e_{\text{att}}(T)\|_{\text{deg}} \tag{53}$$

is also reported, where lower values indicate better landing quality.

Figures 9 and 10 summarize the terminal distributions and Monte Carlo class statistics for the original-disturbance descent scene. This scene corresponds to the trained wind-direction condition for Pure PPO. Under this condition, Pure PPO remains close to the structured INDI, DOB/CTSMC, and SHMPC baselines, with all non-residual controllers achieving mostly Adequate landings and Desired rates near 67–72%. Residual PPO remains higher, reaching a 90.0% Desired rate while preserving 100% Adequate occupancy and giving the smallest composite terminal score.

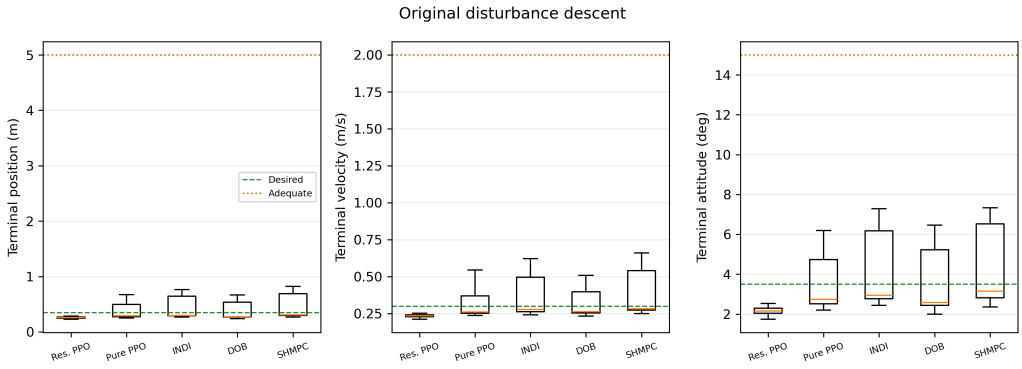

Figure 9. Terminal Monte Carlo distributions for the original disturbance setting extended to descent and landing. Dashed and dotted reference lines indicate the Desired and Adequate limits used for this descent-and-landing test.

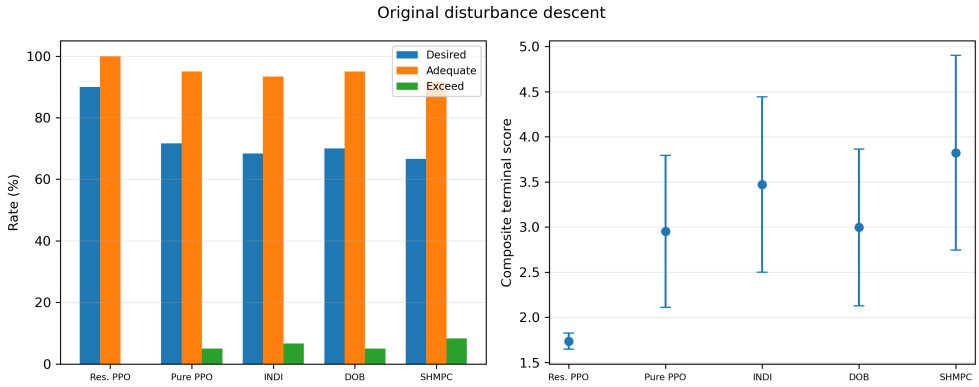

Figure 10. Monte Carlo class rates and terminal-score confidence intervals for the original disturbance setting extended to descent and landing.

Figures 11 and 12 summarize the corresponding statistics for the moving-deck descent scene. This scene is outside the Pure-PPO training condition and therefore exposes its weaker generalization: the Desired rate falls to 25.0%, the Adequate rate falls to 70.0%, and 30.0% of trials exceed the Adequate envelope. In contrast, the structured INDI, DOB/CTSMC, and SHMPC baselines remain mostly Adequate and retain 63.3–68.3% Desired occupancy. Residual PPO achieves the highest Desired rate, 90.0%, while keeping 98.3% of trials within the Adequate envelope.

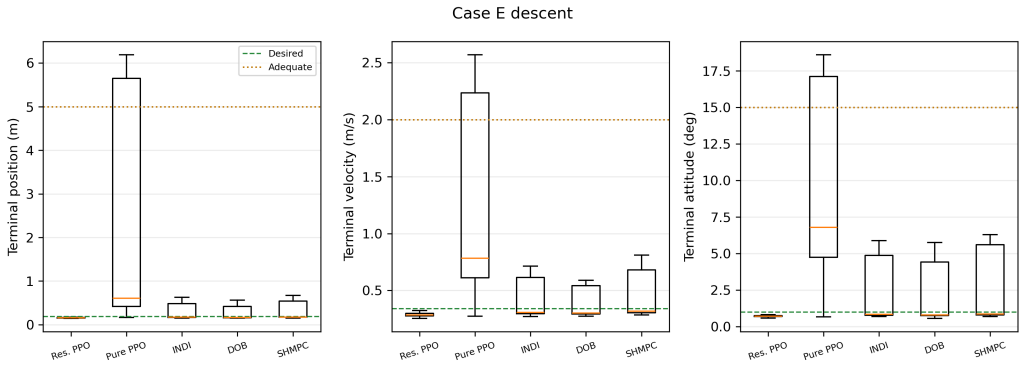

Figure 11. Terminal Monte Carlo distributions for the moving-deck descent scene.

Table 5 summarizes the paired landing Monte Carlo statistics. Each scene uses 60 matched trials with identical initial perturbations across controllers. The table reports Wilson 95% confidence intervals for Desired, Adequate, and exceedance rates, together

with standard deviations and interquartile ranges for continuous metrics. Residual PPO is the only controller that reaches a 90.0% Desired rate in both landing scenes. The pure INDI, DOB/CTSMC, and SHMPC baselines are mostly Adequate and achieve nonzero Desired rates of approximately 63–70%, while Pure PPO is competitive only in the trained original-disturbance scene and degrades substantially in the moving-deck descent scene. Paired statistical comparisons against Residual PPO are reported in Table 6. Binary Desired-rate differences use an exact paired binomial/McNemar test on discordant paired samples, while continuous score differences are tested on matched samples using paired score tests. In the original-disturbance descent scene, Residual PPO improves the Desired rate by 18.3–23.3 percentage points over the structured non-residual controllers. In the moving-deck descent scene, the advantage is 21.7–26.7 percentage points over the structured baselines and 65.0 percentage points over Pure PPO. The terminal-score comparisons also favor Residual PPO in all listed comparisons.

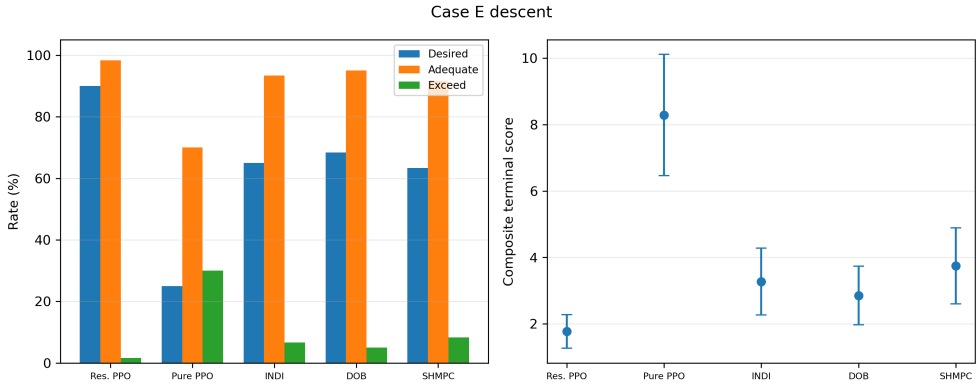

Figure 12. Monte Carlo class rates and terminal-score confidence intervals for the moving-deck descent scene.

Table 5. Two-scene landing Monte Carlo statistics with Wilson 95% confidence intervals for categorical outcomes and dispersion statistics for continuous terminal metrics.

Scene	Controller	Desired (%) [95% CI]	Adequate (%) [95% CI]	Exceed (%) [95% CI]	Position Mean ± SD (m)	Velocity Mean ± SD (m/s)	Score Median [IQR]
Original disturbance descent	Residual PPO	90.0 [79.9, 95.3]	100.0 [94.0, 100.0]	0.0 [0.0, 6.0]	0.271 ± 0.046	0.247 ± 0.041	1.623 [1.597, 1.655]
Original disturbance descent	Pure PPO	71.7 [59.2, 81.5]	95.0 [86.3, 98.3]	5.0 [1.7, 13.7]	0.600 ± 1.093	0.392 ± 0.395	1.861 [1.813, 3.044]
Original disturbance descent	INDI	68.3 [55.8, 78.7]	93.3 [84.1, 97.4]	6.7 [2.6, 15.9]	0.726 ± 1.247	0.455 ± 0.455	1.981 [1.911, 3.934]
Original disturbance descent	DOB/CTSMC	70.0 [57.5, 80.1]	95.0 [86.3, 98.3]	5.0 [1.7, 13.7]	0.602 ± 1.084	0.400 ± 0.418	1.834 [1.784, 3.288]
Original disturbance descent	SHMPC	66.7 [54.1, 77.3]	91.7 [81.9, 96.4]	8.3 [3.6, 18.1]	0.834 ± 1.413	0.499 ± 0.502	2.023 [1.965, 4.122]
Strong-disturbance descent	Residual PPO	90.0 [79.9, 95.3]	98.3 [91.1, 99.7]	1.7 [0.3, 8.9]	0.256 ± 0.656	0.327 ± 0.232	1.449 [1.410, 1.489]
Strong-disturbance descent	Pure PPO	25.0 [15.8, 37.2]	70.0 [57.5, 80.1]	30.0 [19.9, 42.5]	2.070 ± 2.505	1.128 ± 0.828	5.110 [3.868, 18.314]
Strong-disturbance descent	INDI	65.0 [52.4, 75.8]	93.3 [84.1, 97.4]	6.7 [2.6, 15.9]	0.603 ± 1.270	0.513 ± 0.460	1.560 [1.519, 3.963]
Strong-disturbance descent	DOB/CTSMC	68.3 [55.8, 78.7]	95.0 [86.3, 98.3]	5.0 [1.7, 13.7]	0.500 ± 1.131	0.458 ± 0.405	1.521 [1.488, 3.489]
Strong-disturbance descent	SHMPC	63.3 [50.7, 74.4]	91.7 [81.9, 96.4]	8.3 [3.6, 18.1]	0.716 ± 1.437	0.580 ± 0.532	1.613 [1.559, 4.408]

Table 6. Paired statistical tests against Residual PPO for the two landing scenes. Positive Desired-rate effects mean Residual PPO achieved a higher Desired rate; negative score effects mean Residual PPO achieved a smaller terminal score.

Scene	Comparison	Desired-Rate Δpp (p)	Score Δ (p)	Test Basis	Pairs
Original disturbance descent	Residual PPO vs. Pure PPO	+18.3 (<0.001)	−1.2150 (<0.001)	exact binary; paired score	60
Original disturbance descent	Residual PPO vs. INDI	+21.7 (<0.001)	−1.7348 (<0.001)	exact binary; paired score	60
Original disturbance descent	Residual PPO vs. DOB/CTSMC	+20.0 (<0.001)	−1.2598 (<0.001)	exact binary; paired score	60
Original disturbance descent	Residual PPO vs. SHMPC	+23.3 (<0.001)	−2.0873 (<0.001)	exact binary; paired score	60
Strong-disturbance descent	Residual PPO vs. Pure PPO	+65.0 (<0.001)	−6.5218 (<0.001)	exact binary; paired score	60
Strong-disturbance descent	Residual PPO vs. INDI	+25.0 (<0.001)	−1.5008 (<0.001)	exact binary; paired score	60
Strong-disturbance descent	Residual PPO vs. DOB/CTSMC	+21.7 (<0.001)	−1.0821 (<0.001)	exact binary; paired score	60
Strong-disturbance descent	Residual PPO vs. SHMPC	+26.7 (<0.001)	−1.9790 (<0.001)	exact binary; paired score	60

5. Limitations and Future Work

This study is a numerical-simulation-based feasibility investigation. The simulation framework uses simplified modelling assumptions, including a simplified disturbance model, a quasi-steady Cheeseman–Bennett ground-effect correction, an analytic multisine deck-motion model, and a CETI aerodynamic surrogate model. These assumptions are adequate for the comparative tests and algorithm-feasibility verification conducted here, but they cannot fully represent complex real shipboard helicopter operations and should not be interpreted as supporting universal operational conclusions. The descent-and-landing cases reported here should also be interpreted as simplified terminal deck-relative alignment tests; touchdown-contact dynamics, landing-gear interaction, ship-deck contact constraints, and pilot/automation mode transitions are not modeled in detail.

The reported evidence should be read within this tested envelope: 500-run hover Monte Carlo tests at each WOD setting and two 60-pair descent-and-landing studies. Within these tests, Residual PPO achieved 84.8–86.0% Desired hover rates, 95.6–96.6% Adequate hover rates, and 90.0% Desired landing rates in both landing scenes. The residual term was bounded to approximately 20–30% of the nominal outer-loop authority, and the simulations included actuator limits, safety thresholds, and a collective go-around logic. However, formal closed-loop stability proofs, systematic failure-mode grading, redundant backup-control design, and experimental validation remain outside the present scope.

Future work should therefore combine higher-fidelity airwake, dynamic inflow, inclined- and moving-ground, and touchdown-contact models with broader sea-state, gust, deck-motion, out-of-distribution airwake, and actuator-failure campaigns. These extensions should quantify stability margins, failure-mode classes, emergency-logic effectiveness, and residual-authority sensitivity before stronger claims are made for real shipboard landing operations.

6. Conclusions

This paper developed a deck-relative baseline–residual controller for shipboard helicopter near-deck holding and landing. The nominal pathway combines a split-channel INDI outer loop with a reduced-order DI inner loop, while a bounded PPO residual acts in the same physical outer-loop command channels with approximately 20–30% of the nominal authority. The main contribution is therefore the moving-deck baseline–residual control architecture: it preserves an interpretable stabilizing pathway and uses learning as bounded compensation for disturbance-driven mismatch.

The simulation results show consistent gains within the tested strong-disturbance envelope. In the fixed-WOD hover Monte Carlo test at 0°, Residual PPO satisfied the Desired and Adequate criteria in 393/500 and 477/500 trials, corresponding to 78.6% and 95.4%, with a 96.6% criteria-satisfying hold rate, a 0.241 m median terminal position error, and a 1.227 m 95th-percentile position error. Across the 0°, 45°, and 90° WOD sweep, it maintained 84.8–86.0% Desired rates, 95.6–96.6% Adequate rates, 96.2–97.0% success rates, and 0.246–0.263 m XYZ RMS errors. In the two 60-pair descent-and-landing scenes, it reached a 90.0% Desired landing rate in both scenes, with 100.0% and 98.3% Adequate occupancy, and improved Desired rate by 18.3–26.7 percentage points over the structured baselines and by 65.0 percentage points over Pure PPO in the moving-deck scene. These findings support the proposed residual-control formulation for the simulated envelope, while broader sea states, touchdown/contact dynamics, hardware constraints, and formal safety properties still require higher-fidelity and experimental validation.

https://doi.org/10.3390/aerospace13060516

Author Contributions: Conceptualization, X.C. and J.A.; methodology, X.C. and J.A.; software, X.C.; validation, X.C.; formal analysis, X.C.; investigation, X.C.; writing—original draft preparation, X.C.; writing—review and editing, X.C. and J.A.; supervision, J.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The figure, table, and evaluation artifacts supporting the reported results are included in the accompanying submission materials. The simulation code, trained-policy/checkpoint files, detailed configuration files, random seeds, evaluation and plotting scripts, disturbance and deck-motion implementations, and additional intermediate training artifacts are currently being organized for a subsequent public release. Before that release, these materials are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CETI	Control Equivalent Turbulence Input
CB	Cheeseman–Bennett
DI	Dynamic Inversion
DRL	Deep Reinforcement Learning
GE	Ground Effect
INDI	Incremental Nonlinear Dynamic Inversion
MDP	Markov Decision Process
RL	Reinforcement Learning
GAE	Generalized Advantage Estimation
PPO	Proximal Policy Optimization
VF	Value Function
ADS-33E-PRF	Rotorcraft low-speed performance standard referenced in this paper

References

1. Owen, I.; White, M.D.; Padfield, G.D.; Hodge, S.J. A virtual engineering approach to the ship-helicopter dynamic interface: A decade of modelling and simulation research at the University of Liverpool. *Aeronaut. J.* **2017**, *121*, 1833–1857. [CrossRef]
2. Sharma, A.; Padthe, A.; Friedmann, P.P. Helicopter shipboard landing simulation including wind, deck motion and dynamic ground effect. *J. Aircr.* **2021**, *58*, 467–486. [CrossRef]
3. Mehling, T.; Halbe, O.; Gasparac, T.; Vrdoljak, M.; Hajek, M. Piloted simulation of helicopter shipboard recovery with visual and control augmentation. In Proceedings of the AIAA Scitech 2021 Forum, Virtual Event, 11–15, 19–21 January 2021; AIAA Paper 2021-1136. [CrossRef]
4. Lumsden, R.B.; Wilkinson, C.H.; Padfield, G.D. Challenges at the Helicopter-Ship Dynamic Interface. In Proceedings of the 24th European Rotorcraft Forum, Marseilles, France, 15–17 September 1998; Paper OP02.
5. Forrest, J.S.; Owen, I.; Hodge, S.J.; Barakos, G.N.; Padfield, G.D. An investigation of ship airwakes using detached-eddy simulation. *Comput. Fluids* **2010**, *39*, 656–673. [CrossRef]
6. Lee, D.; Sezer-Uzol, N.; Horn, J.F.; Long, L.N. Simulation of helicopter shipboard launch and recovery with time-accurate airwakes. *J. Aircr.* **2005**, *42*, 448–461. [CrossRef]
7. Forrest, J.S.; Owen, I.; Hodge, S.J.; Barakos, G.N.; White, M.D. Evaluating ship superstructure aerodynamics for maritime helicopter operations through CFD and flight simulation. *Aeronaut. J.* **2016**, *120*, 1578–1603. [CrossRef]
8. Topczewski, S.; Narkiewicz, J.; Bibik, P. Helicopter control during landing on a moving confined platform. *IEEE Access* **2020**, *8*, 107315–107325. [CrossRef]
9. Topczewski, S.; Bibik, P. LQR and LQG control of the helicopter during landing on the ship deck. *Aircr. Eng. Aerosp. Technol.* **2023**, *95*, 1344–1352. [CrossRef]
10. Horn, J.F. Non-linear dynamic inversion control design for rotorcraft. *Aerospace* **2019**, *6*, 38. [CrossRef]

11. Simplicio, P.V.M.; Pavel, M.D.; van Kampen, E.-J.; Chu, Q.P. An acceleration measurements-based approach for helicopter nonlinear flight control using incremental nonlinear dynamic inversion. *Control Eng. Pract.* **2013**, *21*, 1065–1077. [CrossRef]
12. Shin, H.; You, D.; Shim, D.H. Autonomous shipboard landing algorithm for unmanned helicopters in crosswind. *J. Intell. Robot. Syst.* **2014**, *74*, 347–361. [CrossRef]
13. Yu, X.; Yang, J.; Li, S. Disturbance observer-based autonomous landing control of unmanned helicopters on moving shipboard. *Nonlinear Dyn.* **2020**, *102*, 131–150. [CrossRef]
14. Huang, Y.; Zhu, M.; Zheng, Z.; Feroskhan, M. Fixed-time autonomous shipboard landing control of a helicopter with external disturbances. *Aerosp. Sci. Technol.* **2019**, *84*, 18–30. [CrossRef]
15. Huang, Y.; Zhu, M.; Zheng, Z. Output-constrained fixed-time control for autonomous ship landing of helicopters. *ISA Trans.* **2020**, *106*, 221–232. [CrossRef]
16. Li, Y.; Huang, Y.; Liu, H.; Li, D. Full state constrained flight tracking control for helicopter systems with disturbances. *Aerospace* **2023**, *10*, 471. [CrossRef]
17. Joelianto, E.; Sumarjono, E.M.; Budiyono, A.; Penggalih, D.R. Model predictive control for autonomous unmanned helicopters. *Aircr. Eng. Aerosp. Technol.* **2011**, *83*, 375–387. [CrossRef]
18. Ngo, T.D.; Sultan, C. Model predictive control for helicopter shipboard operations in the ship airwakes. *J. Guid. Control. Dyn.* **2016**, *39*, 574–589. [CrossRef]
19. Greer, W.B.; Sultan, C. Shrinking horizon model predictive control method for helicopter-ship touchdown. *J. Guid. Control. Dyn.* **2020**, *43*, 884–900. [CrossRef]
20. Ngo, T.D.; Sultan, C. Variable horizon model predictive control for helicopter landing on moving decks. *J. Guid. Control. Dyn.* **2022**, *45*, 774–780. [CrossRef]
21. Johannink, T.; Bahl, S.; Nair, A.; Luo, J.; Kumar, A.; Loskyll, M.; Ojea, J.A.; Solowjow, E.; Levine, S. Residual reinforcement learning for robot control. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*; IEEE: New York, NY, USA, 2019; pp. 6023–6029. [CrossRef]
22. Liu, Y.T.; Price, E.; Black, M.J.; Ahmad, A. Deep residual reinforcement learning based autonomous blimp control. In *Proceedings of the 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*; IEEE: New York, NY, USA, 2022; pp. 12566–12573. [CrossRef]
23. Yang, R.; Du, C.; Zheng, Y.; Gao, H.; Wu, Y.; Fang, T. PPO-based attitude controller design for a tilt-rotor UAV in transition process. *Drones* **2023**, *7*, 499. [CrossRef]
24. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347. [CrossRef]
25. Saetti, U.; Horn, J.F. Flight simulation and control using the Julia language. In Proceedings of the AIAA Scitech 2022 Forum, San Diego, CA, USA, 3–7 January 2022; AIAA Paper 2022-2354. [CrossRef]
26. Howlett, J.J. *UH-60A Black Hawk Engineering Simulation Program: Volume I—Mathematical Model*; NASA Contractor Report CR-166309; NASA: Greenbelt, MD, USA, 1981.
27. Ballin, M.G. *Validation of a Real-Time Engineering Simulation of the UH-60A Helicopter*; NASA Technical Memorandum TM-88360; NASA: Greenbelt, MD, USA,1987.
28. Tischler, M.B. *System Identification Methods for Aircraft Flight Control Development and Validation*; NASA Technical Memorandum TM-110369; USAATCOM Technical Report 95-A-007; NASA Ames Research Center: Moffett Field, CA, USA, 1995.
29. Chen, R.T.N. *A Survey of Nonuniform Inflow Models for Rotorcraft Flight Dynamics and Control Applications*; NASA Technical Memorandum TM-102219; NASA: Greenbelt, MD, USA, 1989.
30. Lusardi, J. Control Equivalent Turbulence Input Model for the UH-60 Helicopter. Ph.D. Thesis, University of California, Davis, CA, USA, 2004.
31. Hess, R.A. A Simplified and Approximate Technique for Scaling Rotorcraft Control Inputs for Turbulence Modeling. *J. Am. Helicopter Soc.* **2004**, *49*, 361–366. [CrossRef]
32. Seher-Weiss, S.; Jones, M. Control equivalent turbulence input models for rotorcraft in hover and forward flight. *J. Guid. Control. Dyn.* **2021**, *44*, 1517–1524. [CrossRef]
33. Ye, Y.; Chen, R.; Ji, H. An active flow control over the ship deck for helicopter shipboard operations. *Aerospace* **2024**, *11*, 119. [CrossRef]
34. Cheeseman, I.C.; Bennett, W.E. *The Effect of the Ground on a Helicopter Rotor in Forward Flight*; Aeronautical Research Council Report and Memorandum No. 3021; Her Majesty's Stationery Office: London, UK, 1957; communicated September 1955.
35. Pasquali, C.; Serafini, J.; Bernardini, G.; Milluzzo, J.; Gennaretti, M. Numerical-experimental correlation of hovering rotor aerodynamics in ground effect. *Aerosp. Sci. Technol.* **2020**, *106*, 106079. [CrossRef]
36. Pasquali, C.; Gennaretti, M.; Bernardini, G.; Serafini, J. State-space dynamic inflow modelling for hovering rotors in fixed-and moving-ground effect. *Aerosp. Sci. Technol.* **2023**, *140*, 108414. [CrossRef]

37. Kang, Q.; Ji, H.; Yuan, Y.; Ye, Y. Autonomous helicopter shipboard recovery flight control design based on tau theory. *Aerosp. Sci. Technol.* **2025**, *159*, 109956. [CrossRef]
38. Schulman, J.; Moritz, P.; Levine, S.; Jordan, M.I.; Abbeel, P. High-Dimensional Continuous Control Using Generalized Advantage Estimation. In Proceedings of the 4th International Conference on Learning Representations (ICLR 2016), San Juan, Puerto Rico, 2–4 May 2016.
39. *ADS-33E-PRF*; Aeronautical Design Standard, Performance Specification: Handling Qualities Requirements for Military Rotorcraft. U.S. Army Aviation and Missile Command: Redstone Arsenal, AL, USA, 2000.
40. Kalra, A.; Maibach, M.-J.; Štrbac, A. An overall evaluation toolchain for assessing helicopter ship deck landings. *CEAS Aeronaut. J.* **2024**, *15*, 425–437. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.