The Aeronautical Journal (2026), 1–25
doi:10.1017/aer.2026.10218

ROYAL AERONAUTICAL SOCIETY

RESEARCH ARTICLE

Learning-based artificial potential field guidance for hypersonic glide vehicle penetration

Xindi Tong[1], Jia Song[1], Wenling Li[2] and Yang Liu[3]

[1]School of Astronautics, Beihang University, Beijing, China
[2]School of Automation Science and Electrical Engineering, Beihang University, China
[3]College of Information Science and Engineering, Northeastern University, China
Corresponding author: Jia Song; Email: songjia@buaa.edu.cn

Received: 11 February 2026 UTC; **Revised:** 11 July 2026 UTC; **Accepted:** 11 July 2026 UTC

Keywords: hypersonic glide vehicle; penetration guidance; twin-delayed deep deterministic policy gradient; artificial potential field method; self-attention mechanism

Abstract

This study investigates the penetration guidance problem for hypersonic glide vehicles in complex constrained environments and proposes a learning-based efficient evasion guidance framework. First, a baseline strategy for determining the vehicle's reference heading in obstacle environments is designed using an artificial potential field method. This baseline algorithm mitigates the inherent contradictions arising from the direct integration of the artificial potential field into a penetration guidance framework. Second, the key sensitive parameter set affecting penetration performance is identified. Based on a deep reinforcement learning (DRL) framework combining the multi-head self-attention mechanism and twin-delayed deep deterministic policy gradient (TD3), an online parameter optimiser dependent on time-series state data is designed. Meanwhile, to ensure the effectiveness of strategy learning for this optimiser and avoid sparse rewards, three reward function models are constructed to guarantee the stable convergence of the agent's strategy learning. This optimiser generates real-time responses for evasion strategy adjustments according to environmental dynamics, thus avoiding the potential limitations of manual offline adjustments and the rigidity of online-loaded strategies. Finally, a series of ablation studies and comparative experiments are conducted. These experiments systematically verify the effectiveness and superior performance of the proposed strategy under both standard and disturbed environmental conditions.

Nomenclature

h (km)	altitude
V (m/s)	Earth-relative velocity
θ (rad)	longitude
ϕ (rad)	latitude
ψ (rad)	heading angle
γ (rad)	flight path angle
α (rad)	angle-of-attack
σ (rad)	bank angle
ψ_{LOS} (rad)	line-of-sight angle
$\Delta\psi$ (rad)	line-of-sight angle error
g (m/s^2)	gravitational acceleration
D (N)	aerodynamic drag force
L (N)	aerodynamic lift force
ρ (kg/m^3)	atmospheric density
m (kg)	mass of the vehicle
S_m (m^2)	reference area

© The Author(s), 2026. Published by Cambridge University Press on behalf of Royal Aeronautical Society.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Q (kW/m^2)	heating rate at a specified point
n (m/s^2)	aerodynamic load
q (Pa)	dynamic pressure
R_e (m)	Earth radius
r (m)	radial distance of the vehicle to the Earth centre
ω_e (rad/s)	Earth angular velocity

1.0 Introduction

Hypersonic glide vehicles (HGVs) have emerged as critical platforms for next-generation aerospace operations, leveraging their wide-speed-regime manoeuvrability and large-scale cross-domain capabilities in near space. Their unique aerothermodynamic adaptation mechanisms and morphing control characteristics have attracted significant research attention [1–3]. However, with the maturation of space-based infrared detection constellations and multi-layered ground-based kinetic interception defenses, these vehicles now face dual suppression from all-time-domain surveillance threats and cross-medium coordinated strikes. Concurrently, the proliferation of avoidance zones comprising geopolitically sensitive areas and multilateral policy buffer regions necessitates real-time autonomous evasion strategy generation. Developing intelligent penetration guidance architectures with environmental responsiveness thus becomes essential to overcome mission constraints imposed by complex flight environments [4].

Under rigidly anchored terminal point constraints, reentry guidance schemes for HGVs bifurcate into two categories: nominal trajectory guidance and predictor-corrector guidance. The core principle of nominal trajectory guidance involves offline design of flight trajectories satisfying all constraints, with mission execution achieved through online trajectory tracking. This approach offloads computational burden, reducing demands on onboard computing resources, yet lacks responsiveness to environmental uncertainties and fails to accommodate state perturbations or deviations [5, 6]. In contrast, predictor-corrector guidance dynamically adjusts flight strategies based on real-time vehicle states and environmental dynamics, thereby achieving both guidance precision and robustness [7–10]. Against the backdrop of rapid advancements in onboard hardware, the predictor-corrector framework emerges as the superior paradigm for developing intelligent penetration algorithms.

The fundamental framework of predictor-corrector comprises longitudinal and lateral guidance links. The lateral guidance link directly regulates the heading by controlling line-of-sight angle error, with its core concept being to determine the expected reference heading and adjust the reversal logic of the bank angle accordingly. This approach is widely adopted by researchers in constructing penetration guidance strategies for reentry vehicles under no-fly-zone (NFZ) constraints. Mature methodologies for determining expected reference headings in NFZ environments fall into two categories. The first establishes reference headings via geometric relationships between NFZs and the vehicle. Serving as reference points for adjusting bank angle reversal logic, NFZs and key waypoints guide the vehicle through lateral guidance cycles [11]. Liang et al. [12] achieved NFZ avoidance by designing expected waypoints, whose vectors incorporate the position and directional constraints of the trajectory. While this method reduces the computational complexity of determining the reference heading, it necessitates precise manoeuvres within a narrower solution space and requires designing rational waypoint strategies. The second category calculates reference headings in real-time using the concept of field. Common approaches include the interfered fluid dynamic system (IFDS) and artificial potential field (APF) methods. For the IFDS-based obstacle avoidance principle, disturbance functions are introduced into an ideal flow field to generate collision-free navigation streamlines [13, 14]. Celestini et al. [15] proposed a 3D trajectory planning strategy combining IFDS with Bézier curves under multiple obstacle constraints. Wu et al. [16] integrated the concept of receding horizon control with IFDS to establish a 3D path planning framework for non-convex obstacle environments. The core principle of APF-based obstacle avoidance involves driving a moving object toward the target while avoiding obstacles through virtual potential field forces, essentially constructing a superimposed field of target-attractive and obstacle-repulsive potentials [17–19]. A lateral guidance algorithm combining APF with velocity azimuth error

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

corridors was designed to reduce heading errors and achieve NFZ avoidance while addressing terminal and path constraints [20]. Hu et al. [21] developed an improved APF method to formulate NFZ judgement criteria and determine reference headings, along with an adaptively updated real-time heading corridor for distributed NFZ constraints. Due to HGVs' long-range and wide-velocity envelope manoeuvring patterns, migrating IFDS-based solutions to HGV penetration strategies carries risks such as disturbance attenuation, global information dependency and computational complexity divergence. In contrast, APF-based approaches are more compatible due to higher computational efficiency, lower dependency on global information and stable behavioural patterns.

However, although the aforementioned schemes demonstrate satisfactory avoidance performance in preset scenarios, it must be acknowledged that critical algorithmic parameters such as potential field functions and weighting coefficients exhibit strong environmental sensitivity. Yet these key parameter sets typically rely on simplified manual adjustment rules with online loading, lacking timely feedback and targeted adjustments based on environmental state awareness throughout mission execution. Therefore, developing efficient sensitive parameter set tuning strategies to enhance the adaptability of vehicle avoidance algorithms is imperative. Scholars have achieved certain progress by developing heuristic algorithm-based optimisers: Wu et al. [22] introduced a method using a deterministic annealing strategy to improve potential field functions through temperature parameter incorporation. A dynamic enhanced firework algorithm was applied to resolve local minima issues in traditional artificial potential field functions, with experimental verification confirming trajectory safety and smoothness enhancements [23]. Nevertheless, heuristic algorithms remain fundamentally stochastic function optimisers dependent on static fitness functions to evaluate solution quality and incapable of perceiving state transitions in dynamic environments. This inevitably causes frequent failures in non-stationary scenarios. Crucially, the high dynamic motion pattern induces intense and frequent state changes in HGVs. Moreover, HGVs impose extremely demanding computational efficiency requirements on optimisers. Given the generally high computational complexity of heuristic algorithms, they represent a suboptimal choice.

As artificial intelligence technology advances at a fast pace, online real-time optimisation frameworks built on deep reinforcement learning (DRL) have stood out as more effective options for adjusting sensitive parameters. These frameworks find wide use in adaptive optimisation, control systems and other relevant fields [24, 25]. Compared to heuristic-based algorithms, DRL functions as an environment-interactive policy optimiser that explicitly models temporal causality between states and actions through Markov decision processes (MDPs), thereby possessing online environmental response capabilities. Additionally, DRL's pre-trained policy networks can rapidly adapt to similar tasks via fine-tuning, significantly reducing cross-scenario deployment costs. Luo et al. [26] introduced a virtual potential field detection zone to enhance the artificial potential field method, integrated with a long short-term memory (LSTM) network-based reinforcement learning agent for online adjustment of detection zone radius to achieve obstacle avoidance for unmanned vehicles. However, when designing online efficient avoidance strategies for HGVs using artificial potential field methods, targeted improvements to conventional approaches must account for both vehicle-specific characteristics and flight environment properties, while ensuring the online self-adaptiveness of modified models.

Inspired by the above literature review, this paper proposes a learning-based APF penetration guidance algorithm for HGVs under multi NFZ constraints, with primary contributions summarised as follows:

(1) Comprehensive consideration of HGV-specific characteristics and flight environment properties establishes an APF benchmark model and its sensitive parameter sets suitable for long-range flight under multiple NFZ constraints. This benchmark model facilitates reference heading design in preset flight scenarios while mitigating conflicts among obstacle avoidance, trajectory smoothing and target reachability.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

(2) To eliminate dependencies on preset parameters and scenarios in the APF benchmark model while enhancing environmental adaptability and robustness, a DRL agent based on the twin-delayed deep deterministic policy gradient (TD3) framework and transformer backbone architecture is designed for online tuning of the benchmark model's sensitive parameter sets. This agent dynamically adjusts each parameter element by capturing contextual relationships within temporal observations to further optimise avoidance effectiveness.

(3) Building upon reference headings jointly generated by the DRL-based sensitive parameter optimiser and the APF benchmark model, lateral guidance cycles determine bank angle reversal logic to achieve vehicle heading correction. When integrated with longitudinally designed numerical predictor-corrector strategies, a learning-based artificial potential field guidance framework for HGV penetration is proposed.

The subsequent content of this paper is organised as follows. Section 2 provides mathematical problem formulation for HGV penetration guidance, encompassing vehicle dynamics and enforced constraints. Section 3 establishes the learning-based artificial potential field penetration guidance framework, including the APF benchmark model with its sensitive parameter sets and the learning-enabled penetration guidance architecture. Section 4 validates strategy efficiency through numerical simulations and provides corresponding analyses. Section 5 concludes the paper.

2.0 Problem formulation

This section systematically describes the flight dynamics of HGV reentry penetration and multiple enforced constraints.

2.1 Reentry dynamics

Given the extensive range and wide velocity envelope of HGVs during reentry glide, the Earth is modeled as a rotating sphere. On this basis, the complete equations of motion are established, in which the first three dynamic equations govern the evolution of the velocity vector (including its magnitude and flow angles), while the remaining three kinematic equations map the vehicle's position within the spherical coordinate system [2, 27, 28]

$$\begin{cases} \dfrac{\mathrm{d}V}{\mathrm{d}t} = -\dfrac{D}{m} - g\sin\gamma + \omega_e^2 r\cos\phi(\sin\gamma\cos\phi \\ \qquad\qquad - \cos\gamma\sin\phi\cos\psi) \\ \dfrac{\mathrm{d}\gamma}{\mathrm{d}t} = \dfrac{1}{V}\left[\dfrac{L\cos\sigma}{m} - g\cos\gamma + \dfrac{V^2\cos\gamma}{r} + 2\omega_e V\cos\phi\sin\psi\right. \\ \qquad\qquad \left. +\omega_e^2 r\cos\phi(\cos\gamma\cos\phi + \sin\gamma\sin\phi\cos\psi)\right] \\ \dfrac{\mathrm{d}\psi}{\mathrm{d}t} = \dfrac{1}{V}\left[\dfrac{L\sin\sigma}{m\cos\gamma} + \dfrac{V^2\cos\gamma\sin\psi\tan\phi}{r}\right. \\ \qquad\qquad -2\omega_e V(\tan\gamma\cos\phi\cos\psi - \sin\phi) \\ \qquad\qquad \left. +\dfrac{\omega_e^2 r}{\cos\gamma}\sin\phi\cos\phi\sin\psi\right] \\ \dfrac{\mathrm{d}r}{\mathrm{d}t} = V\sin\gamma \\ \dfrac{\mathrm{d}\theta}{\mathrm{d}t} = \dfrac{V\cos\gamma\sin\psi}{r\cos\phi} \\ \dfrac{\mathrm{d}\phi}{\mathrm{d}t} = \dfrac{V\cos\gamma\cos\psi}{r} \end{cases} \tag{1}$$

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

$$\begin{cases} L = \frac{1}{2}\rho V^2 C_L S_m \\ D = \frac{1}{2}\rho V^2 C_D S_m \end{cases} \tag{2}$$

where $\rho = \rho_0 \exp(-h/h_s)$ denotes the atmospheric density at flight altitude, ρ_0 is the sea-level atmospheric density and h_s denotes the standard reference altitude.

Considering that aerodynamic characteristics are highly vehicle-dependent, the Common Aero Vehicle (CAV-H) is adopted as the benchmark vehicle in this paper. Consequently, the corresponding aerodynamic coefficients C_L and C_D are formulated as nonlinear functions of the angle-of-attack α and velocity V [28], expressed as:

$$\begin{cases} C_L = -0.2355 + 2.9451\alpha + 0.2949\ \exp(-0.0003943V) \\ C_D = 0.0234 + 2.3795\alpha^2 + 0.3983\ \exp(0.0010794V) \end{cases} \tag{3}$$

2.2 Multiple constraints

To accomplish long-range reentry glide penetration missions, vehicles must satisfy constraint sets including trajectory safety constraints, target reachability constraints, bounded control constraints and penetration constraints.

(1) Trajectory safety constraints

HGVs endure extreme aerothermodynamic environments during reentry glide, necessitating strict control of critical parameters including heating rate $\dot{Q}$, g-load overload n and dynamic pressure q to ensure trajectory safety, mathematically modeled as [7]:

$$\begin{cases} \dot{Q} = K_Q \rho^{0.5} V^{3.15} \le \dot{Q}_{\max} \\ n = \sqrt{D^2 + L^2}/\mathrm{mg} \le n_{\max} \\ q = \frac{1}{2}\rho V^2 \le q_{\max} \end{cases} \tag{4}$$

where $K_Q = 7.9686 \times 10^{-5}$ is a constant.

Additionally, the quasi-equilibrium glide condition (QEGC) [10, 29] is incorporated into trajectory safety constraints as a soft constraint to reduce or prevent trajectory oscillation and bouncing during glide, expressed as:

$$\frac{L\cos\sigma_{EQ}}{m} - g\cos\gamma + \frac{V^2\cos\gamma}{r} \ge 0 \tag{5}$$

where σ_{EQ} represents σ that complies with QEGC.

By employing QEGC, the inequality constraints in Equation (4) can be transformed into bounded constraints on the magnitude of σ. Within these bounds, the constraints in Equation (4) can be considered satisfied.

(2) Target reachability constraints

To precisely quantify terminal guidance precision, the final constraints for reentry flight are mathematically expressed as [4]:

$$\begin{cases} \Delta h = \left|h_t - h_f\right| \le \delta_h \\ \Delta V = \left|V_t - V_f\right| \le \delta_V \\ \Delta L_r = \left|L_{r,t} - L_{r,f}\right| \le \delta_{L_r} \end{cases} \tag{6}$$

where h_t, V_t and $L_{r,t}$ denote altitude, velocity and remaining range upon mission termination, respectively, while h_f, V_f, $L_{r,f}$, δ_h, δ_V, δ_{L_r} represent corresponding desired terminal states and their tolerance thresholds.

The mathematical expression for the remaining range (L_r) to the target is given by:

$$L_r = R_e \arccos[\sin\phi\sin\phi_f + \cos\phi\cos\phi_f\cos(\theta_f - \theta)] \tag{7}$$

where (θ_f, ϕ_f) denotes the longitude and latitude coordinates of the target point.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

(3) Bounded control constraints

Key control variables for reentry guidance include α and σ, both subject to bounded limitations mathematically expressed as:

$$\begin{cases} \alpha_{\min} \le \alpha \le \alpha_{\max} \\ 0 \le |\sigma| \le \sigma_{\max} \end{cases} \tag{8}$$

where $\alpha_{\min}$, $\alpha_{\max}$, and $\sigma_{\max}$ denote the boundary values.

(4) Penetration constraints

Penetration constraints mandate strict avoidance of prohibited airspace during reentry flight, including threat zones and geopolitically sensitive regions. These restricted areas are uniformly treated as NFZs and modeled using standard convex polyhedron modeling as follows [4, 8, 16]:

$$\begin{cases} \Gamma_k(P) = \left[(\theta - \theta_k)/a_k\right]^{2p_k} + \left[(\phi - \phi_k)/b_k\right]^{2q_k} + \left[(\bar{h} - \bar{h}_k)/c_k\right]^{2r_k} \\ P \quad = \left[\theta, \phi, \bar{h}\right]^T \end{cases} \tag{9}$$

where $k = 1,2,3,\ldots,K$ denotes the NFZ index, $\bar{h} = (h + R_e)/R_e$ is the normalised vehicle altitude, and $[\theta_k, \phi_k, \bar{h}_k]^T$ is the centroid coordinates of threat elements. Parameters a_k, b_k, c_k and p_k, q_k, r_k represent radius and shape coefficients of this standard convex polyhedron model.

Equation (9) constitutes a continuously differentiable function monotonically increasing with the distance between adversarial agents. $\Gamma_k(P) \le 1$ indicates the envelope intersection between the penetration vehicle and threat elements, thus the interceptor constraint during adversarial engagement can be defined as:

$$\Gamma_k(P) > 1 \tag{10}$$

Considering NFZs as cylinders of infinite height, the parameters in Equation (9) are configured as follows: $a_k = b_k = R_{N,k}$, $p_k = q_k = 1$, $r_k > 1$ and c_k is assigned a sufficiently large positive constant ($c_k = 1000$ in this study), where $R_{N,k}$ denotes the NFZ radius.

Consequently, vehicles executing reentry penetration missions must avoid all predefined static NFZs while achieving high-precision arrival at the terminal, subject to full enforcement of multi-constraint sets specified in Equations (4)–(8) and (10).

3.0 Learning-based APF guidance

This paper proposes a learning-based APF penetration guidance framework by integrating numerical predictive correction algorithms to address the multi-constraint sets established in Section 2 for reentry glide flight. The proposed framework comprises two core components: a fundamental APF penetration guidance benchmark model and a learning-enabled online optimisation mechanism.

3.1 APF penetration guidance

(1) Longitudinal guidance design

The longitudinal guidance phase requires determining the angle-of-attack and bank angle magnitude. Inspired by the predictor-corrector framework, prior studies typically model reentry glide guidance as a single-variable root-finding problem with respect to bank angle [2, 5, 8, 27, 30]. This architecture determines angle-of-attack commands offline for online loading, alleviating computational burden and enhancing control efficiency.

(a) Angle-of-attack profile design

The predefined angle-of-attack logic addresses two design requirements: first, during initial glide with high altitude and velocity, large angles of attack are implemented to satisfy thermal protection

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

needs and prevent overload in thermal management systems. Second, maximum lift-to-drag ratio angles are adopted to enhance lateral manoeuvrability and extend longitudinal range in mid-late glide phases. Consequently, this section designs the following angle-of-attack control logic:

$$\alpha = \begin{cases} \alpha_{\max}, & V_1 \le V \\ \frac{\alpha_{\max L/D} - \alpha_{\max}}{V_2 - V_1}(V - V_1) + \alpha_{\max}, & V_2 \le V < V_1 \\ \alpha_{\max L/D}, & V < V_2 \end{cases} \tag{11}$$

where $\alpha_{\max L/D}$ denotes the maximum lift-to-drag ratio angle, and V_1, V_2 represent critical parameters governing profile geometry that can be tuned offline based on initial flight conditions.

(b) Bank angle magnitude design

Following completion of angle-of-attack logic design, the bank angle becomes the sole adjustable control variable for online trajectory planning. Simultaneously, based on the defined α profile and incorporating atmospheric density ρ with QEGC from Equation (4), trajectory constraints in Equation (4) are reduced to upper-bound constraints on the magnitude of σ [5, 10]:

$$\begin{cases} \left|\sigma_{\dot{Q}}(V)\right| \le \arccos\left[\left(g - \dfrac{V^2}{r}\right)\dfrac{2mV^{4.3}}{C_L S_m}\left(\dfrac{K_Q}{\dot{Q}_{\max}}\right)^2\right] = \left|\sigma_{\dot{Q}_{\max}}\right| \\ \left|\sigma_n(V)\right| \le \arccos\left[\left(g - \dfrac{V^2}{r}\right)\dfrac{\sqrt{1+(C_D/C_L)^2}}{g n_{\max}}\right] = \left|\sigma_{n_{\max}}\right| \\ \left|\sigma_q(V)\right| \le \arccos\left[\left(g - \dfrac{V^2}{r}\right)\dfrac{m}{C_L q_{\max} S_m}\right] = \left|\sigma_{q_{\max}}\right| \end{cases} \tag{12}$$

Consequently, this study enforces saturation constraints on the bank angle magnitude as shown below to satisfy the constraints in Equation (4).

$$|\sigma_{\max}(V)| = \min\{|\sigma_{\dot{Q}_{\max}}|, |\sigma_{n_{\max}}|, |\sigma_{q_{\max}}|\} \tag{13}$$

The longitudinal guidance task thus transforms into acquiring the required bank angle magnitude under constraints of Equation (13) while incorporating the offline-designed angle-of-attack profile from Equation (11). The numerical predictor-corrector-based longitudinal strategy predicts terminal states during flight and corrects control commands based on deviations from expected states. First, to achieve desired terminal altitude and velocity, the following normalised variables are defined:

$$e = 1/\bar{r} - \bar{V}^2/2 \tag{14}$$

where $\bar{r} = r/R_e$ and $\bar{V} = V/\sqrt{gR_e}$ denote normalised geocentric radial distance and normalised velocity, respectively.

The bank angle magnitude is configured as a piecewise function of energy parameter e:

$$|\sigma(e)| = |\sigma_{\text{opt}}| + (e - e_0)\cdot(\sigma_c - |\sigma_{\text{opt}}|)/(e_f^* - e_0) \tag{15}$$

where σ_c denotes a prespecified constant, $\left|\sigma_{\text{opt}}\right|$ represents the bank angle magnitude solution satisfying requirements via predictive correction, while e_0 and e_f^* correspond to energy states at the current flight condition and the desired terminal state, respectively.

Upon obtaining $\left|\sigma_{\text{opt}}\right|$, the complete bank angle magnitude function is determined. Within each guidance cycle, the solver pseudocode for $\left|\sigma_{\text{opt}}\right|$ appears in Algorithm 1 with c_1 as a predefined constant, c_2, c_3 and c_5 being preselected small positive constants, and designated c_4 as a preset large constant.

The function $f(|\sigma_{\text{opt}}|)$ in Algorithm 1 is defined as follows:

$$f(|\sigma_{\text{opt}}|) = S_{p,\text{togo}} - S_{f,\text{togo}} \tag{16}$$

where $S_{p,\text{togo}}$ and $S_{f,\text{togo}}$ denote the flight range from the vehicle's current position to the predicted terminal and the specified terminal, respectively. $S_{f,\text{togo}}$ can be calculated by Equation (7). The calculation formula

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Algorithm 1: Solving the parameter $|\sigma_{\text{opt}}|$ based on the predictor-corrector method [1].

Input: The flight state vector x and expected target point.
Output: Parameter $|\sigma_{\text{opt}}|$.
Initial the parameters as $|\sigma_{\text{opt}}| = c_1$, $\delta_\sigma = c_2$, $\delta_e = c_3, f(|\sigma_{\text{opt}}|) = c_4$.
Calculate the expected terminal energy e_f^* by Equation (14).
while $f(|\sigma_{\text{opt}}|) > \delta_\sigma$ **do**
  Initial the state x and the predicted terminal energy $e_p = c_5$.
  Get the value of α by Equation (11).
  while $|e_p - e_f^*| > \delta_e$ **do**
    Update predicted state vector x_p by a numerical integration of Equation (1) with α and $|\sigma_{\text{opt}}|$.
    Calculate the predicted terminal energy e_p by Equation (14).
  end while
  Calculate the function value of $f(|\sigma_{\text{opt}}|)$ by Equation (16).
  if $f(|\sigma_{\text{opt}}|) > \delta_\sigma$ **then**
    Update $|\sigma_{\text{opt}}|$ by Equation (18).
  end if
end while
return $|\sigma_{\text{opt}}|$.

for $S_{p,\text{togo}}$ is given below:

$$S_{p,\text{togo}} = \int \frac{R_e V \cos\gamma}{r} dt \tag{17}$$

The function for updating parameter $\left|\sigma_{\text{opt}}\right|$ in Algorithm 1 is configured with reference to the secant method, and is defined as follows:

$$|\sigma_{\text{opt}}|^{n+1} = |\sigma_{\text{opt}}|^n - \frac{f\left(|\sigma_{\text{opt}}|^n\right)}{f\left(|\sigma_{\text{opt}}|^n\right) - f\left(|\sigma_{\text{opt}}|^{n-1}\right)}\left(|\sigma_{\text{opt}}|^n - |\sigma_{\text{opt}}|^{n-1}\right) \tag{18}$$

where $|\sigma_{\text{opt}}|^n$ denotes the n-th iterative value in the solution process.

(2) APF-based lateral guidance benchmark model

The lateral guidance logic, designed to interface with the numerical predictive correction-based longitudinal framework, typically operates on a heading error corridor. During online guidance, bank angle sign reversal logic corrects the deviation between vehicle heading and expected heading within acceptable limits. In standard precision guidance tasks, the expected heading usually aligns with the line-of-sight angle between the vehicle and desired terminal point [27, 31]. However, successful penetration through NFZs while reaching preset terminals necessitates additional strategies to modify the expected heading. Artificial potential field methods represent motion control approaches leveraging virtual potential fields. Core principles involve constructing superimposed fields combining target-attractive and obstacle-repulsive potentials to guide moving objects along negative gradient directions for autonomous obstacle avoidance and target approach [32]. Within mathematical formulations, targets generate attractive potentials while obstacles create repulsive potentials, with the negative gradient of the total potential defining reference motion directions. This computationally efficient concept has been extensively applied in lateral guidance frameworks. To develop a reentry penetration guidance strategy accommodating multi-no-fly-zone constraints via artificial potential field methods, inspired by previous work [8], this section establishes an APF schematic illustrated in Fig. 1, with lateral guidance strategies based on this field detailed below.

For multi NFZ constraints, a criterion incorporating predictive repulsive potential fields and buffer zones is established as Equation (19). HGVs undergo unpowered decelerative glides without active

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

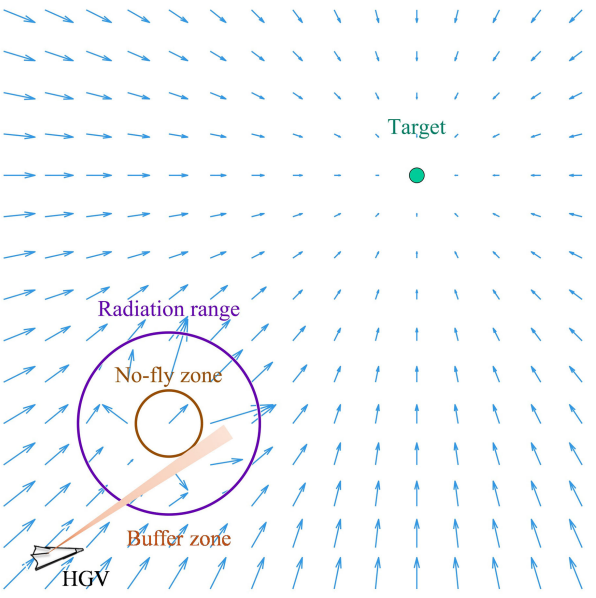


Figure 1. *Adapted APF for highly dynamic reentry flight.*

velocity control. The inability to hover and high velocities necessitate manoeuvre lead margins when designing artificial potential field-based penetration strategies. Predictive repulsive potential fields augment NFZ boundaries with extended influence ranges, activating when vehicles enter these extended zones to modify expected headings. Concurrently, buffer zones resolve directional ambiguity traps occurring when vehicle headings align with NFZ tangent directions, preventing trajectory oscillations at NFZ peripheries caused by evasion strategies.

$$\begin{cases} G^k(t) = \begin{cases} 1, & \text{if } \psi \in \left[\psi_{obs}^k - \zeta_{obs}^k, \psi_{obs}^k + \zeta_{obs}^k\right] \\ 0, & \text{if } \psi > \psi_{obs}^k + \zeta_{obs}^k + \Delta\psi_{buff} \\ & \text{or } \psi < \psi_{obs}^k - \zeta_{obs}^k - \Delta\psi_{buff} \\ G^k(t-1), & \text{otherwise} \end{cases} \\ \psi_{obs}^k = \arctan\left(\dfrac{\sin\Delta\theta}{\cos\phi\tan\phi_k - \sin\phi\cos\Delta\theta}\right) \\ \zeta_{obs}^k = \arcsin\left(\dfrac{1}{R_k}\arccos\Phi\right) \\ \text{where} \begin{cases} \Delta\theta = \theta_k - \theta \\ \Phi = \sin\phi\sin\phi_k + \cos\phi\cos\phi_k\cos\Delta\theta \end{cases} \end{cases} \tag{19}$$

where $G^k(t)$ denotes the predictive repulsive field status of the k-th NFZ at time t, $G^k(t) = 1$ indicates consideration of this NFZ's influence, $G^k(t) = 0$ represents the opposite, and $G^k(t-1)$ reflects the status judgement from the previous cycle. ψ_{obs}^k signifies the bearing angle between the centroid of the k-th NFZ and the vehicle's mass centre, ζ_{obs}^k is the tangent angle between the NFZ's circumference and the vehicle's mass centre, while $\Delta\psi_{buff}$ defines the buffer zone's angular range, encompassing the sector radiating outward from tangents drawn between the vehicle's mass centre toward the NFZ boundary.

A novel exponential-form attractive potential field $U_{gra}(\mathbf{d})$ and its attractive function $\mathbf{F}_{gra}(\mathbf{d})$ are defined as follows. Unlike traditional quadratic potentials, the exponential form alleviates magnitude

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

mismatch between attraction and repulsion over long-range scenarios.

$$U_{gra}(\mathbf{d}) = \frac{1}{2} K_{gra} \exp\left(-||\mathbf{d} - \mathbf{d}_f||^2\right) \tag{20}$$

$$\begin{aligned} \mathbf{F}_{gra}(\mathbf{d}) &= -\nabla U_{gra}(\mathbf{d}) \\ &= \| \mathbf{d} - \mathbf{d}_f \| K_{gra} \exp\left(- \| \mathbf{d} - \mathbf{d}_f \|^2\right) \mathbf{n}_{gra} \end{aligned} \tag{21}$$

where K_{gra} denotes the attractive potential constant, $\mathbf{d} = [\theta, \phi]^T$ and $\mathbf{d}_f = \left[\theta_f, \phi_f\right]^T$ represent current and preset terminal positions, respectively, $||\mathbf{d} - \mathbf{d}_f||$ is the distance vector to the terminal and $\mathbf{n}_{gra}$.

Then, the improved repulsive potential field $U^k_{rep}(\mathbf{d})$ and repulsive force function $F^k_{rep}(\mathbf{d})$ are defined below. When addressing constraints from multiple NFZs, sole generation of attraction by the target may cause excessive repulsive forces leading to failure in reaching the destination. Thus, besides intrinsic repulsive forces F^k_{rep1}, corresponding virtual attractive forces F^k_{rep2} (direction aligned with attraction) are designed for each repulsive field to alleviate inherent conflicts between penetration and target reachability.

$$U^k_{rep}(\mathbf{d}) = \begin{cases} K_\infty, & \text{if } \delta_d < R_k, \mathcal{A} \\ \frac{1}{2} K_{rep} \Omega^2 \|\mathbf{d} - \mathbf{d}_f\|^4, & \text{if } R_k \le \delta_d \le \rho_k, \mathcal{A} \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

$$\mathbf{F}^k_{rep}(\mathbf{d}) = \begin{cases} \mathbf{K}_\infty, & \text{if } \delta_d < R_k, \mathcal{A} \\ F^k_{rep1} \mathbf{n}^k + F^k_{rep2} \mathbf{n}^k_g, & \text{if } R_k \le \delta_d \le \rho_k, \mathcal{A} \\ 0, & \text{otherwise} \end{cases} \tag{23}$$

where $\delta_d = \| \mathbf{d} - \mathbf{d}_k \|$, $\Omega = \delta_d^{-1} - \rho_k^{-1}$, and $\mathcal{A}$ represents the condition $G^k(t) = 1$. The scalar components are given by:

$$F^k_{rep1} = K_{rep} \Omega \| \mathbf{d} - \mathbf{d}_f \|^4 \delta_d^{-2} \tag{24}$$

$$F^k_{rep2} = -2 K_{rep} \Omega^2 \| \mathbf{d} - \mathbf{d}_f \|^3 \tag{25}$$

where K_∞ denote a parameter approaching infinity, K_{rep} is the repulsive potential coefficient, $\mathbf{d}_k = [\theta_k, \phi_k]^T$ indicates the position of the k-th NFZ, $||\mathbf{d} - \mathbf{d}_k||$ is the distance vector between the vehicle and NFZ centre and ρ_k is the influence radius of the k-th NFZ. Unit vectors $\mathbf{n}^k$ and $\mathbf{n}^k_g$ represent directional vectors of F^k_{rep1} and F^k_{rep2}, respectively.

Consequently, the reference heading ψ^* derived from the APF under multi NFZ constraints is:

$$\begin{cases} \mathbf{F}_\Sigma(\mathbf{d}) = \mathbf{F}_{gra}(\mathbf{d}) + \sum_{k=1}^{K} \mathbf{F}^k_{rep}(\mathbf{d}) \\ \psi^* = \angle \mathbf{F}_\Sigma(\mathbf{d}) \end{cases} \tag{26}$$

Finally, the APF-based lateral penetration guidance benchmark model is established as:

$$\begin{cases} \text{sign}(\sigma^t) = \begin{cases} -1, & \psi \ge \psi_{up} \\ 1, & \psi \le \psi_{down} \\ \text{sign}\left(\sigma^{t-1}\right), & \text{otherwise} \end{cases} \\ \psi_{up} = \psi^* + \Delta\psi_{\lim} \\ \psi_{down} = \psi^* - \Delta\psi_{\lim} \end{cases} \tag{27}$$

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

where $\Delta\psi_{\text{lim}}$ denotes the heading error corridor boundary designed as shown in Equation (28):

$$\Delta\psi_{\text{lim}} = \begin{cases} \Delta\psi_1, & V \in [0, V_{\psi,1}] \\ \dfrac{\Delta\psi_2 - \Delta\psi_1}{V_{\psi,2} - V_{\psi,1}}(V - V_{\psi,1}) + \Delta\psi_1, & V \in (V_{\psi,1}, V_{\psi,2}) \\ \Delta\psi_2, & V \in [V_{\psi,2}, \infty) \end{cases} \tag{28}$$

where $\Delta\psi_1$ serve as tunable parameters defining corridor width and inflection points.

Remark 1. The proposed APF-based penetration guidance benchmark model incorporates predictive repulsive fields, buffer zones and novel potential functions to specifically address limitations of conventional artificial potential fields. However, sensitivity to algorithmic parameters persists within the sensitive parameter set: $\Delta\psi_{buff}$ in Equation (19), K_{gra} in Equation (20) and K_{rep}, ρ_k in Equation (22). Conventional approaches rely on offline parameter tuning to enhance adaptability [8, 33], an empirically driven adjustment scheme necessitating extensive experimentation. Considering implementability, such tuning typically yields time-invariant feasible parameters. Nevertheless, diverse flight environments demand varying penetration opportunities and strategies. Particularly under initial state perturbations, fixed parameters inevitably violate boundary constraints or even cause mission failure. Therefore, establishing an online parameter optimisation mechanism that adjusts the sensitive parameter set based on real-time vehicle states and environments is essential to improve the adaptability and robustness of penetration algorithms.

3.2 Learning-based APF guidance strategy

As specified in Remark 1, the sensitive parameter set for the APF-based penetration strategy is denoted by $P_c = \{\Delta\psi_{buff}, K_{gra}, K_{rep}, \rho_k | k = 1,2,\ldots,K\}$. This set critically determines the effectiveness of the guidance strategy across diverse flight environments and mission profiles. The sensitivity analysis regarding the aforementioned parameter ensemble is presented below:

(1) $\Delta\psi_{buff}$. Derived from the specific design of Equation (19), this parameter defines the magnitude of the buffer zone at the tangent of the NFZ. It primarily functions within the joint determination criteria for multiple NFZs, serving as the temporal trigger for incorporating a specific NFZ into the anticipated heading calculation when the aircraft's heading approaches tangency with the NFZ boundary. An excessively large value for this parameter may result in an oversight of the NFZ's influence on the heading, potentially leading to penetration failure. Conversely, an undersized value can cause frequent heading reversals along the tangent direction, inducing trajectory oscillations.

(2) K_{gra},K_{rep}. These two parameters directly determine the magnitudes of attractive and repulsive forces within the APF, serving as critical factors for calculating the anticipated heading. When the attractive potential constant is excessively high, the aircraft prioritises target reachability above all else, which increases the risk of penetration failure. Conversely, if the repulsive potential constant is too large, the aircraft focuses primarily on obstacle avoidance to ensure penetration, potentially leading to target loss. Consequently, only a well-calibrated configuration of these parameters can ensure a rational design of force interaction within the potential field, thereby achieving a balance between target reachability and reliable evasion.

(3) $\rho_k | k = 1,2,\ldots,K$. This parameter determines the influence radius of the NFZ, which uses the actual physical radius as its minimum value. In light of the unpowered flight mode of high-dynamic vehicles during the re-entry glide phase, the design of the anticipated heading strategy must adequately account for the requirement of manoeuvring lead. This requirement is particularly critical during the early stages of flight because the vehicle operates at hypersonic speeds with a large turning radius and limited heading correction capacity. Consequently, a substantial predictive margin is necessary to ensure that the vehicle does not violate the actual NFZ

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

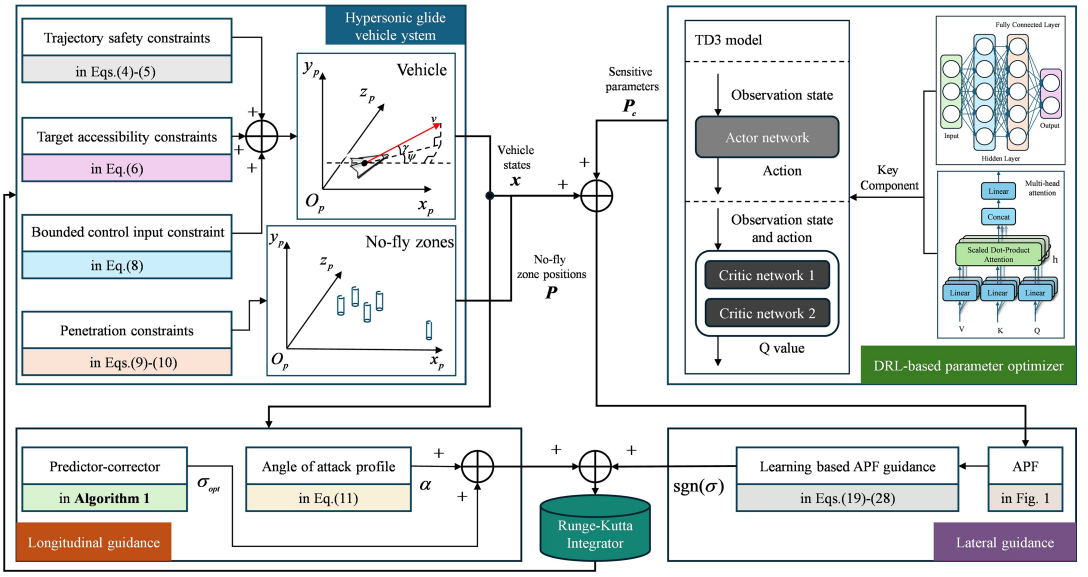


Figure 2. *The learning-based APF guidance strategy architecture.*

boundaries. It should also be noted that expanding the NFZ influence range inevitably narrows the penetration corridors and further compresses the feasible trajectory planning space. Therefore, the adaptive adjustment of the NFZ influence radius can enhance the compatibility and adaptability of the vehicle when it encounters various NFZ deployments across the entire flight profile.

Balancing optimality in parameter tuning and real-time requirements for online guidance, an online parameter optimiser was constructed within a DRL framework. Within a single guidance cycle, this optimiser achieves adaptive adjustments of P_c based on the vehicle's real-time states and surrounding environment. Integrating the mathematical model from Section 2 with the APF-based penetration algorithm in Section 3.1 yields the learning-based APF guidance strategy architecture illustrated in Fig. 2.

Within the proposed learning-based APF penetration guidance framework, a DRL architecture constructs the sensitive parameter optimiser. During each guidance cycle, the optimiser extracts critical features from real-time vehicle states and environmental conditions to determine optimal sensitive parameter sets, constituting a continuous control problem.

The TD3 algorithm demonstrates exceptional adaptability for continuous control tasks, with core advantages stemming from systematic improvements addressing inherent DRL challenges [34]. To accommodate high-dimensional continuous action spaces and precision requirements, TD3 employs twin critic networks to mitigate overestimation bias in value functions. Parallel Q-function estimators operate as:

$$y_1 = r + \gamma \min_{i=1,2} Q_{\theta_i'}(s', \pi_{\varphi_1}(s')) \tag{29}$$

where y_1 denotes the target Q-value, r is the immediate reward and the min operator selects conservative value estimates from twin Q-networks. γ represents the discount factor, while $Q_{\theta_i'}$ corresponds to the target critic network.

Concurrently, delayed policy updates ensure stability in gradient-based policy optimisation, preventing training oscillations caused by frequent policy iterations in conventional methods. Target policy

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

smoothing injects adaptive noise into the action space:

$$\tilde{a} = \pi_{\varphi'}\left(s'\right) + \varepsilon, \quad \varepsilon \sim \mathrm{clip}\left(\mathcal{N}(0, \sigma^2), -c, c\right) \tag{30}$$

where $\tilde{a}$ signifies the noise-injected action, $\pi_{\varphi'}(s')$ is the action generated by the target policy network and ε truncated Gaussian noise with standard deviation σ and clipping range $[-c, c]$. The clip$(\cdot)$ function ensures noise confinement within physically feasible bounds.

This technique enhances robustness against observational noise and model uncertainties while avoiding local optima during parameter search. Notably, convergence reliability improves significantly in non-convex, multi-extremum optimisation scenarios. Symbols in Equations (29) and (30) maintain consistency with the original literature without conflating with other variables in this work.

Additionally, due to the coupling between the sensitive parameter optimiser and guidance cycles, state data involved in decision-making inherently possesses temporal attributes. Thus, this section incorporates multi-head attention mechanisms from transformer architectures atop fully connected layers to capture contextual relationships among state-space variables for decision support. Multi-head attention demonstrates exceptional efficiency in extracting features from sequential data [35]. Its core advantage lies in transcending limitations of conventional sequence models through parallelised multi-perspective decomposition and dynamic weight focusing. Temporally, the multi-head structure decomposes hidden representations of input sequences into subspaces, each independently learning distinct temporal dependency patterns. Certain heads focus on feature interactions within local time windows while others capture long-range dependencies across extended periods. This multi-granularity feature decoupling enables simultaneous identification of high-frequency fluctuations and low-frequency trend evolutions in time series. The fundamental computational model of multi-head attention is given by Equation (31).

$$\begin{cases} \mathrm{Attention}\left(Q, K, V\right) = \mathrm{soft\,max}\left(\dfrac{QK^T}{\sqrt{d_k}}\right)V \\ \mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, \ldots, \mathrm{head}_h)W^O \\ \mathrm{where\ head}_i = \mathrm{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right) \end{cases} \tag{31}$$

where the parameter matrices $W_i^Q \in \mathbb{R}^{d_{\mathrm{model}} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{\mathrm{model}} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{\mathrm{model}} \times d_v}$ and $W^O \in \mathbb{R}^{hd_v \times d_{\mathrm{model}}}$. The number of heads is chosen as $h = 4$, and $d_k = d_v = d_{\mathrm{model}}/h = 128$. The attention function computes a set of queries simultaneously, packing them into matrix Q. Correspondingly, keys and values are packed into matrices K and V, respectively.

In summary, this section designs the agent model for the sensitive parameter optimiser by integrating two key components, fully connected layers and multi-head attention layers, into the TD3-based DRL framework. The training mechanism schematic is shown in Fig. 3, while the structural details of critic and actor networks along with critical parameter designs are specified in Fig. 4.

Following completion of the agent model construction, this section establishes the MDP formulation for the sensitive parameter optimiser as follows.

First, the observation state set must comprehensively encompass the vehicle's current states and relative situational information regarding targets and environmental factors. Sufficient observations provide essential support for the agent's optimisation learning and guide effective decision-making. Therefore, the observation state vector O_t is defined as:

$$\begin{cases} o_t = [r, V, \theta, \phi, \psi, \gamma, \Delta\psi, S_{f,\mathrm{togo}}, \Delta d_{r,1}, \ldots, \Delta d_{r,K}]^T \\ O_t = [o_{t-\tau}, \ldots, o_{t-1}, o_t]^T \end{cases} \tag{32}$$

where o_t denotes the state vector at time t, $r, V, \theta, \phi, \psi, \gamma$ represents the vehicle's flight states, $\Delta\psi, S_{f,\mathrm{togo}}$ indicates positional deviation from the target, $\Delta d_{r,1}, \ldots, \Delta d_{r,K}$ corresponds to relative distances between the vehicle and centroids of NFZs with size $K = 5$, and $\tau = 5$ signifies the temporal dimension of the observation state vector set.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

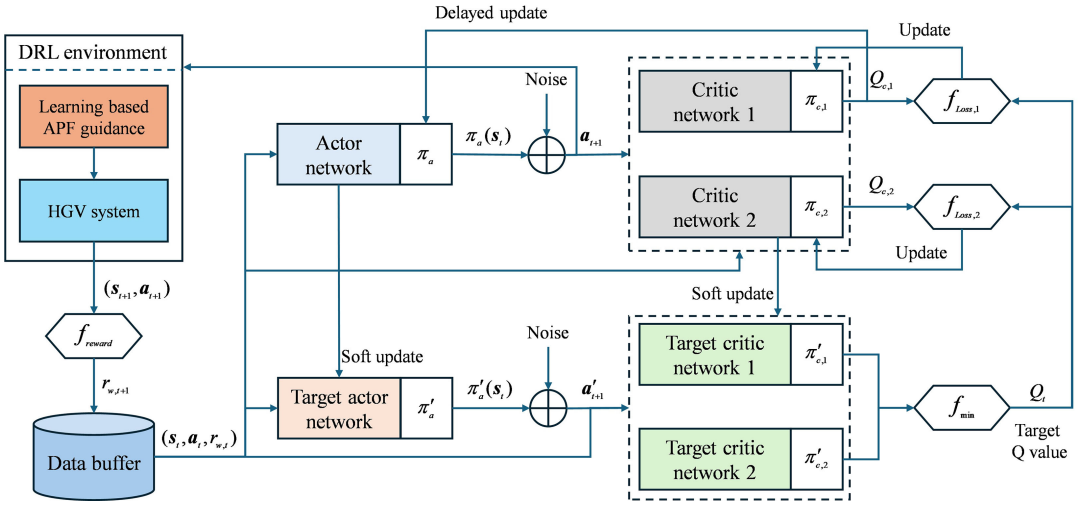


Figure 3. *The training mechanism schematic of the optimiser.*

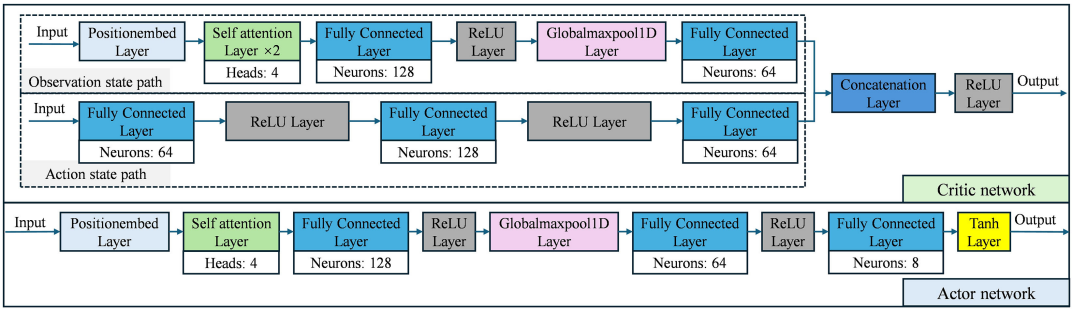


Figure 4. *The network structure and key parameters of the optimiser.*

Remark 2. The underlying design logic and physical insights of the observation vector O_t are elaborated from three aspects to provide comprehensive context for policy optimisation: first, to satisfy the Markov property required for reinforcement learning, the basic motion states $(r, V, \theta, \phi, \psi, \gamma)$ are incorporated to strictly represent the vehicle's instantaneous operational status. Second, to steer the agent to satisfy both terminal guidance objectives and spatial safety constraints, task-oriented features are embedded. Specifically, $\Delta\psi$ and $S_{f,\text{togo}}$ represent the directional and range-to-go errors, offering direct gradients for terminal convergence. Meanwhile, the relative distances $\Delta d_{r,k}$ $(k = 1, \ldots, K)$ map the explicit spatial relationships between the vehicle and multiple NFZs, ensuring that the learned policy can proactively execute avoidance manoeuvres prior to boundary violations. Third, to mitigate the adverse effects of partial observability in complex flight environments, a temporal sequence structure with a history window of $\tau = 5$ is constructed. Stacking these historical observation frames allows the deep neural network to implicitly capture the temporal variations and higher-order derivatives of the vehicle's motion states. This temporal depth significantly stabilises the reinforcement learning training process and enhances the robustness of the derived cooperative guidance law under dynamic environments.

The action space design aligns with the sensitive parameter set as follows:

$$a = \left\{ \Delta\psi_{buff}, K_{gra}, K_{rep}, \rho_k | k = 1,2, \ldots, K \right\} \tag{33}$$

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

where parameters in the action space are bounded within loose boundaries determined through offline testing and analysis. This study selects $\Delta\psi_{buff} \in [2.5^\circ, 7.5^\circ]$, $K_{gra} \in [1 \times 10^3, 9 \times 10^3]$, $K_{rep} \in [5 \times 10^3, 1 \times 10^4]$, $\rho_k \in [1 \times 10^6, 2 \times 10^6]$.

The reward function designed in this study comprises process rewards and terminal rewards, with process rewards consisting of safety reward $R_{s,k}$ and flight reward R_f.

(1) Safety reward serves as the primary feedback to guide the agent toward decisions maintaining safe distances from NFZs, which is critical for fulfilling fundamental avoidance requirements The core principle of $R_{s,k}$ involves regulating the relative distance between the vehicle and NFZ centroids, categorised into danger zones ($\Delta d_{r,k} \le R_{N,k}$), strong penalty zones ($R_{N,k} < \Delta d_{r,k} \le 1.1R_{N,k}$), alert zones ($1.1R_{N,k} < \Delta d_{r,k} \le 1.2R_{N,k}$) and safe zones. Gradient-based designs corresponding to each category guide rational decision-making: (a) danger zones impose substantial fixed penalties and terminate episodes; (b) strong penalty zones apply exponential penalties to drive rapid evasion; (c) alert zones enforce linear penalties to promote gradual distancing; (d) safe zones incur no penalties.

(2) Flight reward R_f relates to spherical distance between the vehicle and target terminal along with flight duration, ensuring adherence to target reachability constraints while avoiding sparsity issues inherent in terminal-only rewards. R_f primarily governs flight time and remaining range, incentivising target-approaching strategies.

(3) Terminal reward R_{goal} provides one-time rewards based on whether preset terminal states are achieved upon episode completion.

Mathematical formulations of these reward functions are defined below:

$$R_{s,k} = \begin{cases} -\lambda_1, & \text{if } \tilde{d}_k \le 1 \\ -\lambda_2 \exp\left(\lambda_3(1-\tilde{d}_k)\right), & \text{if } 1 < \tilde{d}_k \le 1.1 \\ -\lambda_4(1.2-\tilde{d}_k), & \text{if } 1.1 < \tilde{d}_k \le 1.2 \\ 0, & \text{otherwise} \end{cases} \tag{34}$$

where $\tilde{d}_k = \Delta d_{r,k}/R_{N,k}$ denotes the normalised distance.

$$R_f = -\lambda_5(S_{f,\text{togo}} - L_{r,f})/(t \cdot L_{r,f}) \tag{35}$$

$$R_{\text{goal}} = \begin{cases} \lambda_6, & \text{if Eq. (6) is satisfied} \\ -\lambda_7, & \text{otherwise} \end{cases} \tag{36}$$

where $\lambda_1 - \lambda_7$ denote configurable parameters, and t represents flight time.

Consequently, the total reward guiding agent policy optimisation is:

$$R_w = \sum_{k=1}^{K} R_{s,k} + R_f + R_{\text{goal}} \tag{37}$$

Remark 3. Reward functions in Equations (34)–(36) require rational parameter configurations to ensure efficient environmental responses and dynamic multi-objective balance. First, mathematical continuity at segment boundaries of the safety reward must be preserved to avoid abrupt reward changes during boundary transitions. However, stepwise penalties remain permissible for catastrophic failures to deter unacceptable behaviours. Second, magnitude coordination among reward functions is essential for balanced multi-objective optimisation. Based on these considerations and real-time adjustments from training feedback, this study selects: $\lambda_1 = 25$, $\lambda_2 = 22.5$, $\lambda_3 = 30$, $\lambda_4 = 11.25$, $\lambda_5 = 0.025$, $\lambda_6 = 5000$, $\lambda_7 = 10,000$.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Table 1. *Key parameter specifications for agent training*

Parameters	Settings
Observation space dimension	13×5
Action space dimension	8×1
Learning rate of actor	$1e-3$
Learning rate of critic	$2e-4$
L2 regularisation factor of actor	$1e-5$
L2 regularisation factor of critic	$2e-4$
Optimiser	Adam
Attention head count	4
Target smooth factor	$5e-3$
Gradient threshold	1
Discount factor	0.99
Variance of target policy smooth model	0.2
Lower limit of target policy smooth model	−0.5
Upper limit of target policy smooth model	0.5

Table 2. *Configurations for joint NFZs*

NFZs	#1	#2	#3	#4	#5
Position (°E, °N)	(37.5, 6.0)	(32.5, 0.0)	(25.0, 4.0)	(45.0, 2.0)	(52.5, 5.0)
Radius R_N (km)	225	225	225	225	50

Table 3. *Initial and terminal state settings of HGV*

h_0 (km)	V_0 (m/s)	θ_0 (°)	ϕ_0 (°)	ψ_0 (°)	γ_0 (°)	h_f (km)	V_f (m/s)	θ_f (°)	ϕ_f (°)	L_{rf} (km)
60	7000	0	0	90	0	27	1550	55	5	100

Following agent model construction and MDP formulation for the sensitive parameter optimiser, policy learning proceeds according to the Fig. 3 training mechanism with key parameters configured in Table 1.

In summary, the learning-based artificial potential field guidance strategy for HGV penetration has been established.

4.0 Numerical simulations

To validate the efficacy of the learning-based artificial potential field guidance strategy established in Section 3 for penetration scenarios, a series of numerical simulations were conducted using MATLAB 2023b on a computational platform (AMD Ryzen 7 4800H, 2.9 GHz). Detailed parameter configurations, visualised experimental results, and corresponding analyses are provided to ensure simulation reliability and reproducibility.

First, the penetration scenario incorporates five NFZs with deployment positions and influence radii configured as specified in Table 2. Second, rigidly constrained trajectory endpoints for the HGV ***mission*** are defined in Table 3. Finally, parameters governing the penetration guidance strategy in Section 3 are detailed in Table 4.

4.1 Agent training for sensitive parameter optimisation

Building upon the agent architecture, reward mechanisms and training framework established in Section 3, this section conducts agent training for sensitive parameter optimisation. The offline training

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Table 4. *Parameter configuration for the guidance strategy*

ρ_0 (kg/m ^{3})	g (m/s ^{2})	h_s (km)	$\dot{Q}_{\max}$ (kW/m ^{2})	$n_{\max}$ (g)	$q_{\max}$ (kPa)	$\alpha_{\max}$ (°)
1.225	9.8	7.11	6000	4	40	25
$\alpha_{\max L/D}$ (°)	V_1 (m/s)	V_2 (m/s)	$\Delta\psi_1$ (°)	$\Delta\psi_2$ (°)	$V_{\psi,1}$ (m/s)	$V_{\psi,2}$ (m/s)
15	5000	3000	5	12.5	3000	5000

Table 5. *Uniform distributions for aerodynamic uncertainties and initial state dispersions*

h_0 (km)	V_0 (m/s)	θ_0 (°)	ϕ_0 (°)	ψ_0 (°)	γ_0 (°)	m_0 (kg)	C_L –	C_D –	ρ –
±1	±50	±1	±1	±0.5	±0.1	±5%	±10%	±10%	±10%

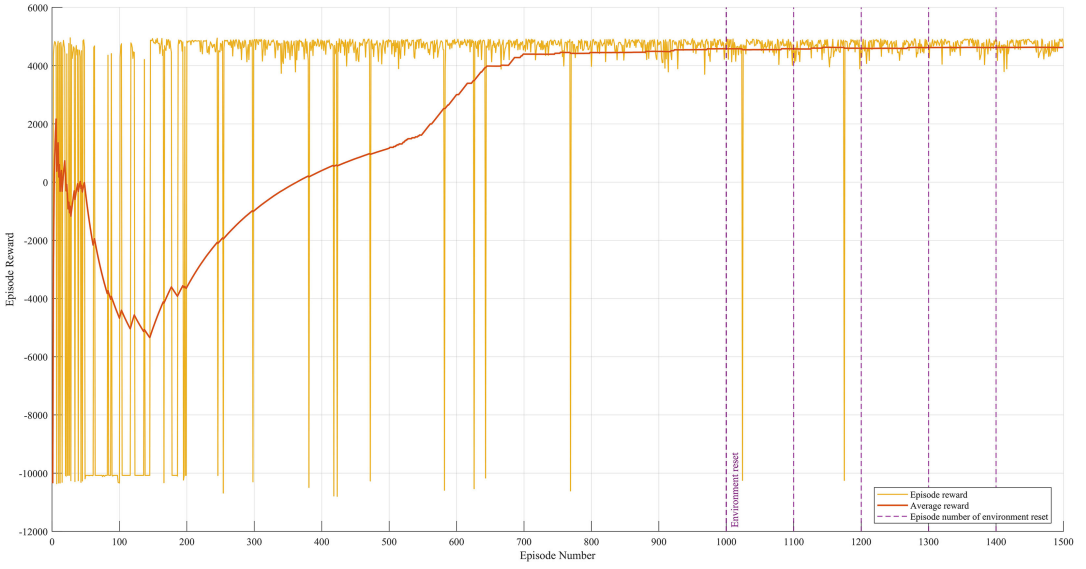
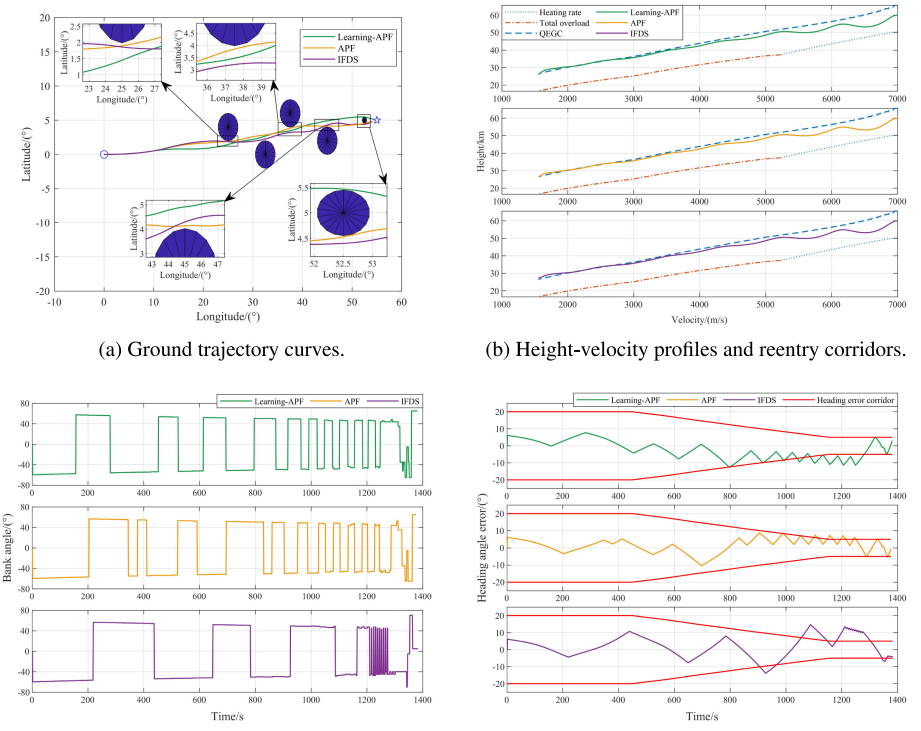


Figure 5. *Reward curves for agent training in sensitive parameter optimisation.*

incorporates comprehensive uncertainties and dispersion errors inherent to HGV systems to enhance the generalisation capability of the DRL method and ensure operational effectiveness of the trained actor network. As detailed in Table 5, aerodynamic model uncertainties and initial state dispersion errors follow uniform distributions calibrated to actual flight system characteristics.

The model training encompasses 1500 episodes. To accelerate exploration of efficient decision policies during initial training phases and mitigate early-stage instability, environment states are initialised with fixed random seeds until significant convergence trends emerge in reward curves. This strategy provides a controlled starting point while reducing stochastic interference during exploratory phases. Upon observing definitive convergence trends in later training stages, the environment initialisation shifts to randomised reset mode, employing dynamically varying random seeds. This introduces environmental uncertainties to evaluate and enhance agent robustness and adaptability under noisy or perturbed conditions. The episodic reward and its moving average throughout training are depicted in Fig. 5.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press



(a) Ground trajectory curves.

(b) Height-velocity profiles and reentry corridors.

(c) Bank angle curves.

(d) Heading error curves and boundaries.

Figure 6. *Comparative simulation results.*

Figure 5 distinctly reveals the agent's learning dynamics. During the initial training phase (first 1000 episodes), the agent operates in an exploration-dominated period. At this stage, limited policy experience results in highly stochastic behaviours that frequently trigger environmental penalties, causing substantial episodic reward fluctuations with pronounced oscillations. As training progresses and policy networks undergo iterative optimisation, the agent progressively learns decision patterns that maximise cumulative rewards by identifying and avoiding hazardous states while pursuing high-reward actions. This capability to construct effective internal representations of environmental dynamics and generate adaptive responses remains robust after introducing dynamically randomised environment resets in later training stages (post-1000 episodes). Confronting abrupt environmental uncertainties and state-space perturbations, the agent rapidly adjusts strategies while maintaining decision robustness and efficiency. Therefore, the final performance demonstrates: (a) significantly reduced fluctuation of episodic rewards with converging distribution stability; (b) progressive improvement in mean reward culminating in statistically significant convergence.

These systemic transformations collectively demonstrate successful acquisition of a stable, adaptive and efficient policy resilient to dynamically evolving environments.

4.2 Effectiveness validation of learning-based APF guidance strategy

To verify the effectiveness and superiority of the proposed learning-based APF guidance strategy in practical flight missions, this section conducts both ablation studies and comparative simulation experiments under standard conditions. The algorithm employed for the ablation study is the APF guidance strategy configured with a fixed sensitive parameter set. To ensure a fair comparison, this parameter set has undergone rigorous offline tuning to ensure its fundamental suitability for the specified flight mission. Furthermore, to evaluate the performance of the proposed algorithm relative to other mainstream evasive guidance methods, the IFDS algorithm is selected as the benchmark based on a comprehensive

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Table 6. *Terminal states in the comparative simulations*

Algorithm	h_t/km	V_t/(m/s)	$L_{r,t}$/km	t_c/s
Learning-APF	26.13 (−3.22%)	1554.70 (+0.30%)	100.56 (+0.56%)	2.014
APF	26.38 (−2.30%)	1556.37 (+0.41%)	101.35 (+1.35%)	1.215
IFDS	27.41 (+1.52%)	1553.23 (+0.21%)	101.24 (+1.24%)	0.816

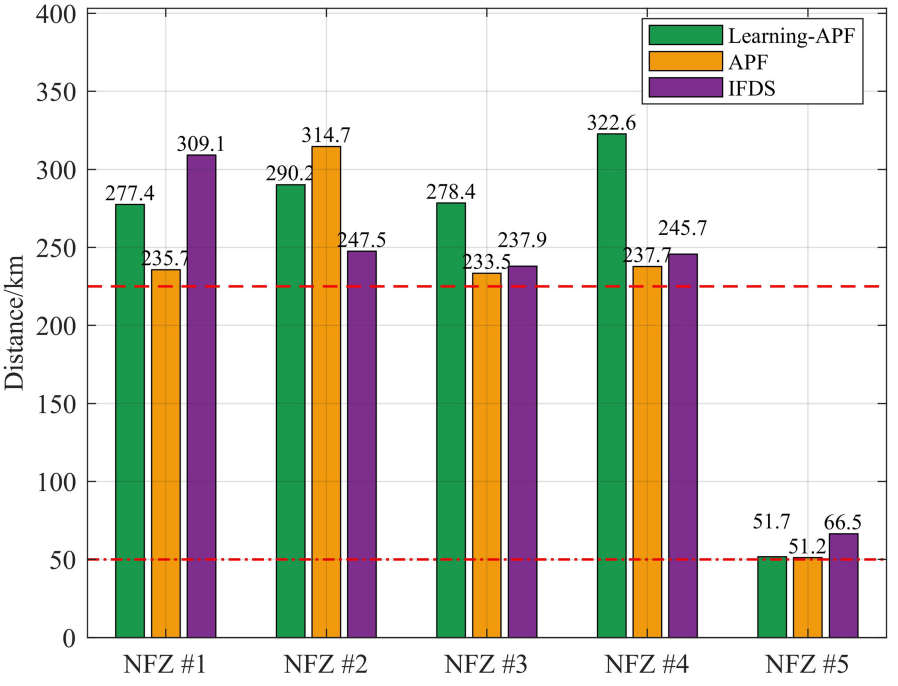


Figure 7. *The minimum distance between the vehicle and NFZs.*

literature review. Since both the IFDS and the APF belong to the category of field-theory-based algorithms, they adhere to the principle of fairness for comparative analysis. Finally, the performance of these three algorithms is analysed and compared based on multiple evaluation metrics. The key parameters for constructing the flight mission environment are detailed in Tables 2 and 3, while the configurable parameters involved in the aforementioned algorithms are specified in Table 4. The simulation results are presented in Figs. 6 and 7 and Table 6, and the detailed analysis and corresponding conclusions are provided as follows:

(1) Figure 6a illustrates the ground trajectories calculated by the three guidance algorithms. Combined with the statistical analysis of the minimum relative distance between the vehicle and the NFZs presented in Fig. 7, it is evident that the vehicles guided by all three algorithms are capable of penetrating the joint NFZ blockade and reaching the designated terminal point. However, due to the distinct intrinsic attributes and parameter configurations of each algorithm, the generated evasive strategies exhibit significant differences under identical environmental constraints.

(2) The curves in Fig. 6b indicate that all three algorithms achieve successful penetration guidance while satisfying the re-entry flight corridor constraints defined in Equation (4). According to EqUationss (12) and (13), the flight corridor constraints composed of process constraints are transformed into amplitude constraints on the bank angle within the longitudinal guidance cycle.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Since these three algorithms employ distinct lateral guidance strategies yet share the same longitudinal guidance framework, these results fully demonstrate the robustness of the longitudinal guidance design proposed in this study.

(3) Figure 6c illustrates the bank angle commands generated by the three algorithms, which are directly correlated with their respective evasive strategies. Specifically, the bank angle manoeuvre pattern of the learning-based APF guidance strategy achieves safe penetration with fewer reversals, thereby further ensuring the stability of the flight trajectory. Moreover, the IFDS algorithm exhibits severe oscillations in heading correction, which further validates the necessity of the buffer zone setting and its rational parameter configuration. Adopting the contact with the NFZ boundary as the sole correction criterion leads to a rapid correction-hold-correction cycle once the vehicle approaches the tangent direction of the NFZ, continuing until the vehicle exits the influence range of the NFZ.

(4) To provide a clearer depiction of the heading correction details for the three guidance algorithms, Fig. 6d presents the heading angle error curves under the lateral heading error corridor constraints. It is worth noting that this error corridor functions as an elastic soft constraint rather than a rigid hard boundary; its primary purpose is to provide corrective regulation whenever a large deviation occurs between the heading angle and the line-of-sight direction to the target. During the midcourse flight, transient excursions beyond this preset corridor are acceptable, as the vehicle must balance the trade-off between two competing objectives, namely obstacle avoidance and terminal reachability. Different avoidance manoeuvres inherently yield varying heading profiles, which merely reflect the distinct manoeuvring strategies adopted by each algorithm. Figure 6d clearly demonstrates that all three methods rigorously fulfill the final terminal heading requirements. Nevertheless, it should be emphasised that both APF-based algorithms are capable of maintaining relatively linear and gradual heading error profiles. As a comparison, the guidance strategy based on IFDS suffers from noticeable fluctuations in heading tracking, as shown in Fig. 6c. Such behaviour reveals that the IFDS framework struggles to balance obstacle avoidance against destination reachability, inducing undesirable trajectory oscillations and compromising the system's safety margin. In contrast, the algorithm developed in this task achieves the desired guidance commands with fewer heading corrections, ensuring higher trajectory stability.

(5) Table 6 presents the performance of the three algorithms in terms of the terminal constraint set and execution time. Since the evasive guidance strategy prioritises the handling of path constraints during flight, and all three algorithms share the same fundamental guidance framework, they exhibit similarly high precision in satisfying terminal constraints. It should be noted that the learning-based APF guidance strategy entails a higher computational overhead due to the requirements of network model loading and forward propagation. Compared with the IFDS algorithm, the proposed APF requires more nested decision-making and related calculations rather than simple physics-based heading computations, which is due to several specific designs tailored for compatibility with the re-entry flight process. Consequently, this leads to a longer solution time. Nevertheless, on the premise of ensuring trajectory reliability and safety, the overall time cost has not increased significantly. Specifically, it should be clarified that the time statistics presented in Table 6 denote the cumulative computational duration required for the complete trajectory generation. For this specific trajectory, the average single-step execution time of the proposed algorithm is calculated as approximately 1.458 ms. According to the state-of-the-art intelligent entry guidance framework [36], the authoritative real-time execution limit for onboard flight computers is established at 50 ms per control cycle (corresponding to a 20 Hz command update frequency). Manifestly, the single-step latency of the proposed strategy (1.458 ms) is well within this required 50 ms ceiling, completely satisfying the onboard real-time standard with an ample safety margin.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

Table 7. *Terminal statistical results of successful cases in Monte Carlo experiments*

Algorithm	h_t (%)	V_t (%)	$L_{r,t}$ (%)	p_{suc} (%)
Learning-APF	2.64 ± 1.33	0.21 ± 0.16	0.77 ± 0.46	99.2
APF	2.71 ± 1.59	0.21 ± 0.15	0.80 ± 0.43	76.2
IFDS	6.58 ± 0.88	0.47 ± 0.14	0.82 ± 0.46	65.6

Table 8. *Comparative results of policy inference efficiency*

Algorithm	Total memory (MB)	Peak memory (MB)	Inference time (s)
Learning-APF	125.151 ± 11.746	4.012 ± 0.001	2.059 ± 0.104
APF	24.149 ± 5.675	2.484 ± 0.166	1.280 ± 0.026
IFDS	27.371 ± 4.868	2.505 ± 0.002	0.829 ± 0.022

4.3 Monte Carlo results

To further verify the anti-disturbance capability and adaptability of the learning-based APF guidance strategy in an uncertain environment, and to simultaneously highlight the advantages of the sensitive parameter optimiser over fixed parameter sets under disturbances to environmental and mission parameters through comparison, this section conducts 500 independent and repeated shooting simulation experiments for each of the three algorithms. The experiments refer to the parameter settings in Table 5 and control the random number seeds within the range of 1–500. The statistical results of the simulations are presented in Figs. 8 and 9 and Tables 7 and 8, and the detailed analysis and corresponding conclusions are provided as follows:

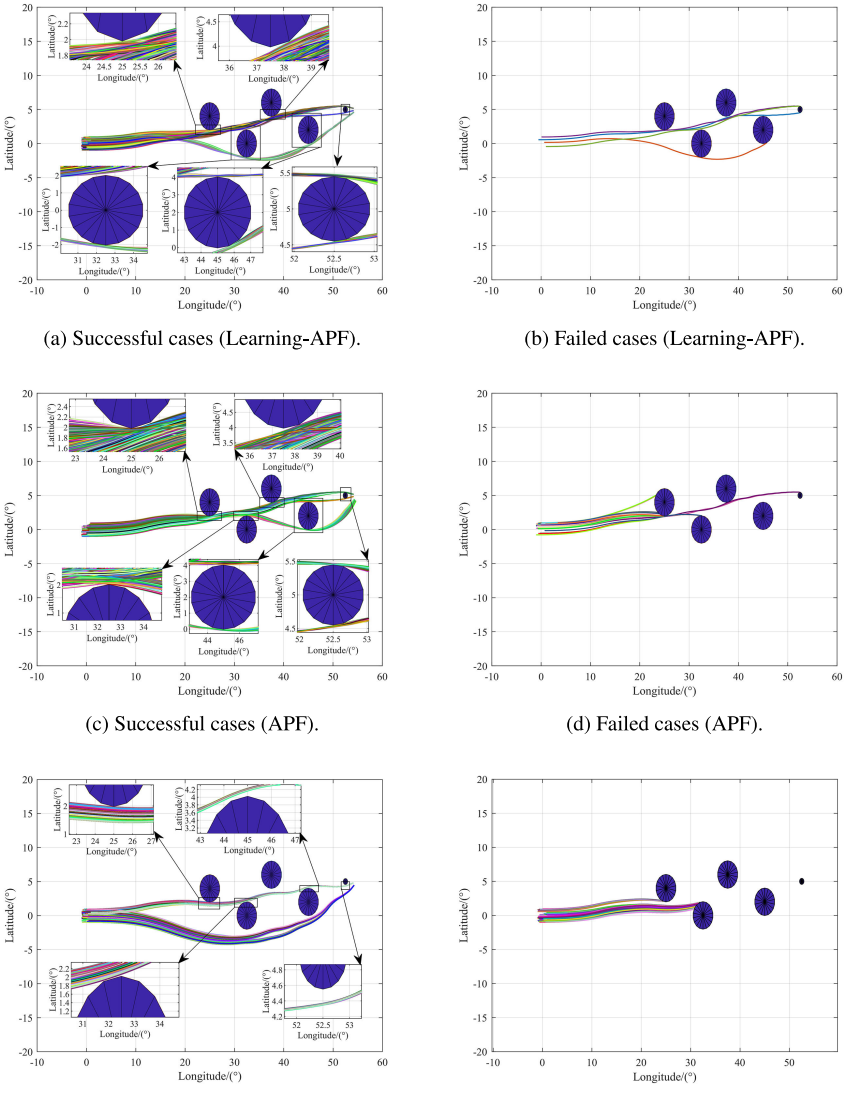
(1) Analysis of Fig. 8a, b and e reveals that even under identical environmental settings, the strategies employed by the three algorithms to achieve successful penetration differ significantly. The learning-based APF strategy flexibly selects evasive manoeuvres according to the vehicle's proximity to NFZs, which significantly enhances the smoothness of the generated trajectories. In contrast, the APF algorithm with fixed parameters and the IFDS algorithm exhibit more rigid planning behaviours, as their fixed configurations fail to generate effective adaptive responses when encountering environmental disturbances.

(2) As illustrated in Fig. 8b, d and f, the learning-based APF outperforms the other two algorithms with the lowest number of failure cases. The primary obstacles leading to failures in both the fixed APF and IFDS are the two NFZs encountered during the early flight stage. This phenomenon is attributed to the hypersonic speeds, large turning radii and limited heading adjustment capacity during this period, which necessitate a higher redundancy margin in the evasive strategy. However, the fixed APF cannot adaptively adjust its strategy based on early flight requirements, while the IFDS, relying solely on physical constraints, neglects the specific demands of the re-entry process.

(3) To provide a clearer representation of the performance differences in Monte Carlo simulations, Fig. 9 summarises the penetration success rates for the joint NFZs. It must be acknowledged that the learning-based APF does not achieve a 100% success rate, as shown in Fig. 8b. This is closely related to factors such as the inherent capacity boundaries of the algorithm and the need for further refinement in parameter optimisation. Nevertheless, the learning-based APF demonstrates superior robustness against environmental disturbances. Under the same experimental conditions, the proposed strategy increases the success rate from 76.2% (fixed APF) to 99.2%, significantly exceeding the 65.6% success rate of the IFDS.

(4) The terminal control precision for all successful cases in the Monte Carlo simulations is summarised in Table 7. The data indicate that although the three algorithms share the same basic

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press



(a) Successful cases (Learning-APF).

(b) Failed cases (Learning-APF).

(c) Successful cases (APF).

(d) Failed cases (APF).

(e) Successful cases (IFDS).

(f) Failed cases (IFDS).

Figure 8. *Ground trajectories of the Monte Carlo experiments.*

guidance framework, the lateral errors caused by different evasive strategies must be compensated for within the longitudinal guidance loop. The flexibility of the learning-based APF further releases the control margin for longitudinal guidance, thereby achieving higher terminal control precision.

(5) To further investigate the computational costs on hardware platforms, the memory consumption (including total and peak consumption) and execution time for all test cases are recorded. The analysis of Table 8 shows that compared to the purely mathematical APF and IFDS algorithms, the learning-based APF incurs higher memory and time costs. These are necessary overheads resulting from agent model loading and network forward propagation. Nevertheless, these costs remain well within the operational load requirements of general hardware development platforms.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

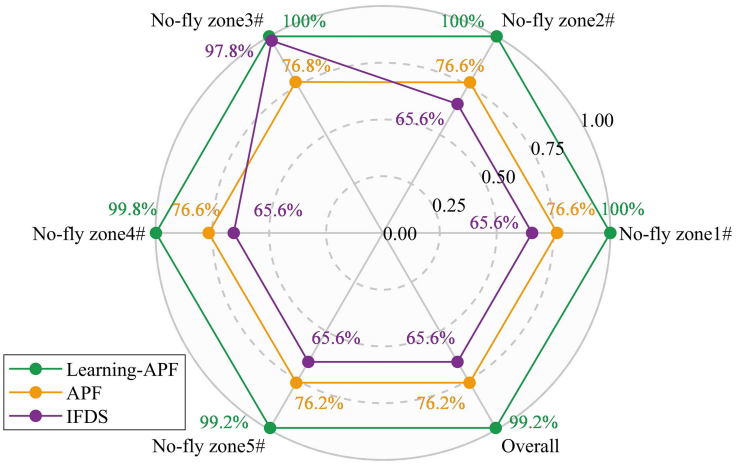


Figure 9. *Comparative results of penetration probability.*

In summary, the learning-based APF guidance strategy proposed in this study can achieve a better balance between target reachability and penetration. It can also significantly improve the rigidity of penetration strategies caused by fixed parameters, all while not compromising terminal control precision or operation time. Therefore, it serves as a more efficient and robust penetration strategy.

5.0 Conclusion

This study addresses the penetration guidance challenge of HGVs in joint NFZs under complex constrained environments, with key conclusions summarised as follows: first, the APF method effectively resolves the inherent contradictions of integrating traditional artificial potential field methods into the HGV penetration guidance framework. It establishes a reasonable reference heading mechanism for obstacle environments, laying a technical foundation for subsequent optimisation and providing a reliable comparison benchmark. Second, the online parameter optimiser is designed based on the multi-head self-attention mechanism and TD3 DRL framework, and it exhibits strong adaptive adjustment capabilities. By leveraging time-series state data to perceive environmental dynamics, the optimiser generates real-time evasion strategy responses. This overcomes the limitations of manual offline adjustment and the rigidity of fixed online strategies. Additionally, the three designed reward function models avoid sparse rewards in reinforcement learning, ensuring stable convergence of the agent's strategy learning. Third, ablation studies and comparative experiments confirm the effectiveness and robustness of the proposed learning-based APF guidance framework. Under both standard and environmental disturbance conditions, the framework maintains high terminal control precision and operational efficiency. It also balances target reachability with penetration success, and significantly enhances strategy flexibility and adaptability to uncertain environments compared to fixed-parameter strategies. In conclusion, this learning-based evasion guidance framework offers a new technical approach for improving HGVs' penetration capability in complex environments. However, a potential limitation lies in the framework's generalisation boundaries under extreme, unmodeled environmental anomalies, and the continuous parameter adaptation may incur additional computational overhead. Future research may expand environmental disturbance scenarios or optimise the reinforcement learning network structure to enhance real-time performance and generalisation for more challenging flight conditions.

Acknowledgements. This work was supported in part by the Key Program of National Natural Science Foundation of China under Grant U24B20177, in part by State Key Laboratory of High-Efficiency Reusable Aerospace Transportation Technology, and in part by Beijing Key Laboratory of System Design for Reusable Launch Vehicle.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

References

[1] Shen, Z., Yu, J., Dong, X., Hua, Y. and Ren, Z. Penetration trajectory optimization for the hypersonic gliding vehicle encountering two interceptors, *Aerosp. Sci. Technol.*, 2022, **121**, pp 107363.

[2] Tong, X., Song, J., Xu, C. and Yu, J. Two-stage spatiotemporal cooperative reentry guidance strategy using transformer and improved beluga whale optimization, *Control Eng. Practice*, 2024, **153**, pp 106078.

[3] Yu, X., Li, P. and Zhang, Y. Fixed-time actuator fault accommodation applied to hypersonic gliding vehicles, *IEEE J. Robot. Autom.*, 2021, **18**, (3), pp 1429–1440.

[4] Wu, T., Wang, H., Liu, Y., Li, T. and Yu, Y. Learning-based interfered fluid avoidance guidance for hypersonic reentry vehicles with multiple constraints, *ISA Trans.*, 2023, **139**, pp 291–307.

[5] Wang, X., Guo, J., Tang, S., Qi, S. and Wang, Z. Entry trajectory planning with terminal full states constraints and multiple geographic constraints, *Aerosp. Sci. Technol.*, 2019, **84**, pp 620–631.

[6] Zhang, Y., Chen, K., Liu, L., Tang, G. and Bao, W. Entry trajectory planning based on three-dimensional acceleration profile guidance, *Aerosp. Sci. Technol.*, 2016, **48**, pp 131–139.

[7] Lu, P. Entry guidance: a unified method, *J. Guid. Control Dyn.*, 2014, **37**, (3), pp 713–728.

[8] Tong, X., Song, J., Li, W. and Xu, C. Penetration game strategy of high dynamic vehicles with constraints of No-fly zones and interceptors, *Eng. Appl. Artif. Intell.*, 2024, **136**, p 109018.

[9] Wang, T., Zhang, H., Zeng, L. and Tang, G. A robust predictor–corrector entry guidance, *Aerosp. Sci. Technol.*, 2017, **66**, pp 103–111.

[10] Xue, S. and Lu, P. Constrained predictor-corrector entry guidance, *J. Guid. Control Dyn.*, 2010, **33**, (4), pp 1273–1281.

[11] Wang, T., Zhang, H. and Tang, G. Predictor-corrector entry guidance with waypoint and no-fly zone constraints, *Acta Astronaut.*, 2017, **138**, pp 10–18.

[12] Liang, Z., Liu, S., Li, Q. and Ren, Z. Lateral entry guidance with no-fly zone constraint, *Aerosp. Sci. Technol.*, 2017, **60**, pp 39–47.

[13] Tamanakijprasart, K., Mondal, S. and Tsourdos, A. Dynamic path planning of UAV in three-dimensional complex environment based on interfered fluid dynamical system, AIAA SciTech Forum Expos., American Institute of Aeronautics and Astronautics Inc., 2024.

[14] Yao, P., Wang, H. and Su, Z. Cooperative path planning with applications to target tracking and obstacle avoidance for multi-UAVs, *Aerosp. Sci. Technol.*, 2016, **54**, pp 10–22.

[15] Celestini, D., Primatesta, S. and Capello, E. Trajectory planning for uavs based on interfered fluid dynamical system and Bézier Curves, *IEEE Robot. Autom. Lett.*, 2022, **7**, (4), pp 9620–9626.

[16] Wu, J., Wang, H., Zhang, M. and Yu, Y. On obstacle avoidance path planning in unknown 3D environments: a fluid-based framework, *ISA Trans.*, 2021, **111**, pp 249–264.

[17] Warren, C.W. Global path planning using artificial potential fields, 1989 IEEE International Conference on Robotics and Automation, 1989, pp 316–321.

[18] Bounini, F., Gingras, D., Pollart, H. and Gruyer, D. Modified artificial potential field method for online path planning applications, IEEE Intelligent Vehicles Symposium Proceedings, 2017, pp 180–185.

[19] Duhé, J.F., Victor, S. and Melchior, P. Contributions on artificial potential field method for effective obstacle avoidance, *Fract. Calc. Appl. Anal.*, 2021, **24**, pp 421–446.

[20] Zhang, D., Liu, L. and Wang, Y. On-line reentry guidance algorithm with both path and no-fly zone constraints, *Acta Astronaut.*, 2015, **117**, pp 243–253.

[21] Hu, Y., Gao, C., Li, J., Jing, W. and Chen, W. A novel adaptive lateral reentry guidance algorithm with complex distributed no-fly zones constraints, *Chin. J. Aeronaut.*, 2022, **35**, (7), pp 128–143.

[22] Wu, Z., Dai, J., Jiang, B. and Karimi, H.R. Robot path planning based on artificial potential field with deterministic annealing, *ISA Trans.*, 2023, **138**, pp 74–87.

[23] Li, H., Liu, W., Yang, C., Wang, W., Qie, T. and Xiang, C. An optimization-based path planning approach for autonomous vehicles using the DynEFWA-artificial potential field, *IEEE Trans. Intell. Veh.*, 2022, **7**, (2), pp 263–272.

[24] Yin, B., Fang, X. and Wang, X. Joint optimization of trajectory control, resource allocation, and user association based on DRL for multi-fixed-wing UAV networks, *IEEE Trans. Wireless Commun.*, 2024, **23**, (10), pp 13330–13343.

[25] Tong, X., Song, J. and Li, W. Rapid decision-making strategy for UAV swarms in complex adversarial environments using proximal policy optimization and transformer, *IEEE Trans. Aerosp. Electron. Syst.*, 2025, **61**, (5), pp 13183–13201.

[26] Luo, J., Wang, Z.X. and Pan, K.L. Reliable path planning algorithm based on improved artificial potential field method, *IEEE Access*, 2022, **10**, pp 108276–108284.

[27] Yu, J., Dong, X., Li, Q., Ren, Z. and Lv, J. Cooperative guidance strategy for multiple hypersonic gliding vehicles system, *Chin. J. Aeronaut.*, 2020, **33**, (3), pp 990–1005.

[28] Xu, H., Cai, G., Mu, C. and Li, X. Analytical reentry guidance framework based on swarm intelligence optimization and altitude-energy profile, *Chin. J. Aeronaut.*, 2023, **36**, (12), pp 336–348.

[29] Guo, Y., Li, X., Zhang, H., Wang, L. and Cai, M. Entry guidance with terminal time control based on quasi-equilibrium glide condition, *IEEE Trans. Aerosp. Electron. Syst.*, 2020, **56**, (2), pp 887–896.

[30] Sana, K.S. and Hu, W. Hypersonic reentry trajectory planning by using hybrid fractional-order particle swarm optimization and gravitational search algorithm, *Chin. J. Aeronaut.*, 2021, **34**, (1), pp 50–67.

[31] Song, J., Tong, X., Xu, X. and Zhao, K. A real-time reentry guidance method for hypersonic vehicles based on a time2vec and transformer network, *Aerospace*, 2022, **9**, (8), pp 427.

[32] Khatib, O. Real-time obstacle avoidance for manipulators and mobile robots, *Proceedings of 1985 IEEE International Conference on Robotics and Automation*, vol. 2, 1985, pp 500–505.

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press

[33] Song, J., Xu, X., Tong, X. and Zhao, K. A time cooperation guidance for multi-hypersonic vehicles based on LSTM network and improved artificial potential field method, *Aerospace*, 2022, **9**, (10), pp 562.
[34] Fujimoto, S., van Hoof, H. and Meger, D. Addressing function approximation error in actor-critic methods, *Proceedings of the 35th International Conference on Machine Learning*, vol. 80, 2018, pp 1587–1596.
[35] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L. and Polosukhin, I. Attention is all you need, Advances in Neural Information Processing Systems 30, 2017.
[36] Cheng, L., Jiang, F., Wang, Z. and Li, J. Multiconstrained real-time entry guidance using deep neural networks, *IEEE Trans. Aerosp. Electron. Syst.*, 2021, **57**, (1), pp 325–340.

Cite this article: Tong X., Song J., Li W. and Liu Y. Learning-based artificial potential field guidance for hypersonic glide vehicle penetration. *The Aeronautical Journal*, https://doi.org/10.1017/aer.2026.10218

https://doi.org/10.1017/aer.2026.10218 Published online by Cambridge University Press