Journal of Electrotechnology, Electrical Engineering and Management (2026)
Clausius Scientific Press, Canada
DOI: 10.23977/jeeem.2026.090104
ISSN 2560-6697 Vol. 9 Num. 1

High-Precision Trajectory Tracking for UAV Aerial Refueling Based on Deep Reinforcement Learning and Adaptive Model Predictive Control

Yanning Gong*

Jiangxi University of Water Resources and Electric Power, Nanchang, Jiangxi, 330099, China
**Corresponding author*

Keywords: Aerial refueling; Trajectory tracking; Model Predictive Control; Deep reinforcement learning; Wake vortex disturbance; Soft Actor-Critic (SAC)

Abstract: To address the challenges of strong wake vortex aerodynamic disturbances, highly nonlinear dynamics of the hose-drogue system, and strict constraints during the unmanned aerial vehicle (UAV) aerial refueling docking phase, a cascaded trajectory tracking and autonomous decision-making framework is proposed, integrating Soft Actor-Critic (SAC) deep reinforcement learning (DRL) with Adaptive Model Predictive Control (AMPC). First, a six-degree-of-freedom (6-DOF) nonlinear dynamic model of the UAV incorporating the Burnham-Hallock wake vortex velocity field interference is established, alongside a lumped-mass catenary dynamic model for the hose-drogue system. Second, at the decision-making layer, an autonomous rendezvous agent based on the SAC algorithm is designed. By formulating a Markov Decision Process (MDP) featuring a continuous state-action space and a composite reward function, safe obstacle avoidance and optimal trajectory planning for the UAV in complex airflow environments are achieved. At the control layer, specifically targeting the precise docking phase, an AMPC is synthesized to enforce physical hard constraints on the actuators (i.e., deflection magnitude and rate limits). Furthermore, a first-order low-pass incremental state observer is introduced to estimate and compensate for the time-varying wind field disturbances induced by the wake vortex in real time. Finally, rigorous numerical evaluations and dynamic simulations are conducted for verification. The results demonstrate that, compared with traditional fuzzy PID and Nonlinear Dynamic Inversion (NDI) control, the proposed SAC-AMPC cascaded framework reduces the Root Mean Square Error (RMSE) along the X, Y, and Z axes by 61.9%, 61.1%, and 64.4%, respectively, relative to the fuzzy PID baseline. Additionally, the maximum transient overshoot is bounded within 0.55 m, the control energy consumption is reduced by 28.8%, and actuator saturation is completely eradicated. The docking success rate across 100 Monte Carlo runs reaches 98.0%. This research provides a novel theoretical foundation and highly reliable technical support for the autonomous aerial refueling of UAVs in complex, disturbance-rich environments.

1. Introduction

Aerial refueling technology is a crucial force multiplier for modern air forces, capable of significantly extending the endurance of unmanned aerial vehicles (UAVs) and enhancing strategic deterrent capabilities. With the accelerated evolution of unmanned combat systems, the "unmanned receiver-manned tanker" and "unmanned-unmanned" refueling modes have emerged as core technologies prioritized by leading military powers worldwide.

The docking phase of UAV aerial refueling is widely recognized as one of the most challenging tasks in the field of flight control. As the receiver UAV approaches the tanker, it must not only satisfy relative pose constraints with extremely tight tolerances (typically $\leq$0.3 m) but also overcome the strong wake vortex disturbances generated by the tanker's wings, atmospheric turbulence, and the nonlinear "whipping" effect of the hose-drogue system induced by the airflow. Traditional control methods exhibit obvious limitations when dealing with such strongly coupled, constrained, and time-varying complex systems. Literature [1] applied fuzzy PID control to achieve the docking of a quadrotor UAV; however, its fuzzy rules rely heavily on expert experience and suffer from phase lag when facing high-frequency wake vortex disturbances, which easily triggers control overshoot. Literature [2] utilized Nonlinear Dynamic Inversion (NDI) for trajectory tracking, but its practical performance is highly dependent on the accuracy of the aerodynamic model, exhibiting weak robustness against wind field disturbances. Literature [3] enhanced the anti-disturbance capability via an LQR-PI controller but failed to explicitly handle the physical hard constraints of the actuators (such as control surface deflection saturation) at the algorithmic level, rendering the system vulnerable to instability under abrupt crosswinds.

In recent years, deep reinforcement learning (DRL) has demonstrated immense potential in autonomous decision-making due to its powerful online exploration and nonlinear mapping capabilities. Meanwhile, Model Predictive Control (MPC) possesses inherent advantages in solving high-precision tracking problems for Multi-Input Multi-Output (MIMO) systems with constraints. Motivated by these facts, this paper proposes an innovative SAC-AMPC cascaded control architecture:

1) To address the issue of state oscillation in traditional finite state machine logic (e.g., SECA rules) under sudden complex airflows, an SAC agent characterized by maximum entropy is designed to output an optimal approach trajectory with foresight and obstacle avoidance capabilities.

2) To solve the challenge of precise docking control, an Adaptive Model Predictive Controller (AMPC) is synthesized. The maneuvering commands are optimized over a rolling prediction horizon, and an incremental observer in the outer loop is utilized to estimate and compensate for the wake vortex interference in real time. Under the premise of strictly satisfying the hard constraints of actuator deflection, extreme precision and rapid docking without overshoot are achieved.

2. Problem Formulation and System Modeling

2.1 UAV 6-DOF Dynamic Model Considering Wake Vortex Interference

Assuming the receiver UAV is a rigid body and neglecting the effects of Earth's rotation and curvature, the nonlinear six-degree-of-freedom (6-DOF) dynamic equations in the body coordinate frame can be expressed as:

$$\begin{cases} \dot{\mathbf{V}}=\frac{1}{m}(\mathbf{F}_{aero}+\mathbf{F}_{thrust}+\mathbf{F}_{gravity})-\boldsymbol{\Omega}\times\mathbf{V} \\ \dot{\boldsymbol{\Omega}}=\mathbf{I}^{-1}(\mathbf{M}_{aero}+\mathbf{M}_{thrust}-\boldsymbol{\Omega}\times(\mathbf{I}\boldsymbol{\Omega})) \\ \dot{\mathbf{P}}=\mathbf{R}_b^e\mathbf{V} \\ \dot{\boldsymbol{\Theta}}=\mathbf{S}(\boldsymbol{\Theta})\boldsymbol{\Omega} \end{cases}$$

where $\mathbf{V}=[u,v,w]^T$ is the body velocity vector; $\boldsymbol{\Omega}=[p,q,r]^T$ represents the body angular velocity; $\mathbf{P}=[x,y,z]^T$ denotes the position in the inertial frame; $\boldsymbol{\Theta}=[\phi,\theta,\psi]^T$ signifies the Euler angles; and $\mathbf{I}$ is the inertia matrix. To reflect the severity of the actual refueling environment, the aerodynamic force $\mathbf{F}_{aero}$ is further decoupled into nominal aerodynamic forces and wake vortex interference forces. The Burnham-Hallock wake vortex velocity field model is introduced[6]:

$$V_\theta(r)=\frac{\Gamma_0}{2\pi r}\frac{r^2}{r^2+r_c^2}$$

where r is the radial distance from the receiver UAV to the center of the wake vortex, r_c is the vortex core radius, and Γ_0 denotes the vortex circulation strength. The actual airspeed of the receiver UAV is superimposed with the downwash and sidewash velocity components induced by the wake vortex, constituting a strongly time-varying wind field disturbance $\mathbf{V}_{wake}$.

2.2 Hose-Drogue Catenary Dynamic Model

During the aerial docking process, the drogue is not a static target; instead, it undergoes a "whipping" oscillation due to the combined effects of atmospheric turbulence and the tanker's wake vortex. The lumped-mass method is adopted to discretize the hose into N rigid links, with the drogue treated as the $(N+1)$-th point mass. The differential equation of motion for the i-th point mass is given by:

$$m_i\ddot{\mathbf{P}}_i=\mathbf{T}_i-\mathbf{T}_{i-1}+\mathbf{G}_i+\frac{1}{2}\rho V_a^2 S_i C_D\frac{\mathbf{V}_a}{//\mathbf{V}_a//}$$

where $\mathbf{T}_i$ is the tension vector between adjacent links. The dynamic three-dimensional coordinates of the terminal point mass $\mathbf{P}_{drogue}(t)$ serve as the time-varying tracking target for the receiver UAV's controller.

3. Autonomous Decision-Making and Trajectory Planning Based on SAC Reinforcement Learning

To enable the receiver UAV to smoothly and safely transition from the observation area to the docking corridor without being drawn into the core wake vortex region of the tanker, the Soft Actor-Critic (SAC) algorithm[5] is employed for high-level autonomous decision-making. The "maximum entropy" characteristic of SAC effectively prevents the policy from prematurely converging to local suboptimal solutions.

3.1 State and Action Space Design

The **State Space (S_t)** is defined by the relative kinematic information between the receiver UAV and the drogue:

$$S_t=[\Delta x,\Delta y,\Delta z,\Delta u,\Delta v,\Delta w,\phi,\theta,\psi]^T$$

The **Action Space (A_t)** is designated as the desired trajectory angular rate and acceleration commands required by the lower-level controller:

$$A_t=[\dot{\psi}_{cmd},\dot{\theta}_{cmd},a_{cmd}]^T$$

3.2 Construction of the Composite Constraint Reward Function

To balance promptness, smoothness, and safety, a composite reward function incorporating guiding and penalizing terms is designed:

$$R_t=-c_1 // \mathbf{P}_{uav}-\mathbf{P}_{drogue} //_2+c_2\exp(-c_3|\Delta\psi|)-c_4 // A_t-A_{t-1} //_2+R_{safe}$$

where the first two terms reward distance reduction and heading alignment; the third term penalizes abrupt changes in actions to guarantee flying qualities; R_{safe} acts as a hard penalty term that assigns a massive negative constant and terminates the episode if the receiver UAV breaches the collision envelope or deviates from the docking corridor.

4. High-Precision Trajectory Tracking and Anti-disturbance Control Based on AMPC

When the SAC guides the receiver UAV into the precise docking zone (distance to the drogue ≤10 m), the system switches to the AMPC controller. During this phase, the system must precisely handle the physical saturation of the actuators and rapidly suppress the positional deviations caused by the high-frequency swinging of the drogue.

4.1 Linearized Predictive Model and Adaptive Observer

The nonlinear model is subjected to local Jacobian linearization at the current equilibrium point and discretized using Euler's method to obtain the augmented state-space model:

$$\begin{cases}\mathbf{x}(k+1)=\mathbf{A}_k\mathbf{x}(k)+\mathbf{B}_k\mathbf{u}(k)+\mathbf{B}_d\hat{\mathbf{d}}(k)\\ \mathbf{y}(k)=\mathbf{C}_k\mathbf{x}(k)\end{cases}$$

where the state vector $\mathbf{x}$ comprises position, velocity, and attitude errors; the control input $\mathbf{u}=[\delta_e,\delta_a,\delta_r,\delta_{th}]^T$ corresponds to the elevator, aileron, rudder, and throttle commands, respectively. To address unmodeled aerodynamic characteristics and wake vortex disturbances, an adaptive incremental observer is designed[4]:

$$\hat{\mathbf{d}}(k)=\lambda\hat{\mathbf{d}}(k-1)+(1-\lambda)[\mathbf{x}(k)-(\mathbf{A}_k\mathbf{x}(k-1)+\mathbf{B}_k\mathbf{u}(k-1))]$$

4.2 Constrained Rolling Optimization Problem (QP) Solving

Within the prediction horizon N_p and control horizon N_c, a cost function considering the restrictions on control increments is constructed:

$$\min_{\Delta\mathbf{U}}\mathbf{J}=\sum_{i=1}^{N_p} // \mathbf{y}(k+i|k)-\mathbf{y}_{ref}(k+i|k) //_{\mathbf{Q}}^2+\sum_{j=0}^{N_c-1} // \Delta\mathbf{u}(k+j|k) //_{\mathbf{R}}^2$$

The constraints include the **absolute deflection limits** and **maximum deflection rate limits** of the actuators:

$$\begin{cases}\mathbf{u}_{min}\le\mathbf{u}(k+j|k)\le\mathbf{u}_{max}\\ -\Delta\mathbf{u}_{max}\le\Delta\mathbf{u}(k+j|k)\le\Delta\mathbf{u}_{max}\end{cases}$$

In each control cycle, the Interior Point Method is utilized to solve this Quadratic Programming (QP) problem online, and the first element of the optimal control sequence $\mathbf{u}^*(k)$ is dispatched to the actuators. This fundamentally circumvents the loss of control caused by integrator windup, a phenomenon frequently encountered in traditional PID controllers.

5. Simulation Experiments and Result Analysis

To verify the robustness and accuracy of the proposed SAC-AMPC cascaded control framework in highly nonlinear and intensely disturbed environments, a semi-physical simulation verification platform was constructed based on the Robot Operating System (ROS) and the JSBSim 6-DOF flight dynamics engine. The tanker was set as a large fixed-wing platform maintaining level flight at a constant altitude of 6000 m and a speed of 120 m/s. The receiver UAV initiated rendezvous decision-making via the SAC agent from an observation area (50 m behind and below the target drogue) and switched to the AMPC for precise docking during the final 10 meters.

5.1 Core Simulation Parameters and Environment Construction

The Burnham-Hallock wake vortex disturbance and a moderate-intensity Dryden gust model were injected into the simulation environment. Strict physical saturation limits (magnitude and deflection rate limits) were applied to the aileron, elevator, and rudder of the receiver UAV. The core dynamics and algorithmic parameters are detailed in Table 1.

Table 1 Core Parameter Settings for Simulation Experiments

Parameter Category	Parameter Name	Value and Unit
UAV Physical Parameters	Receiver mass m	2500 kg
	Maximum thrust T_{max}	8000 N
	Deflection limits $[\delta_{min},\delta_{max}]$	$[-25^\circ,25^\circ]$
	Maximum deflection rate $\dot{\delta}_{max}$	$60^\circ/s$
Wake Vortex & Environment	Vortex circulation Γ_0	450 m^2/s
	Vortex core radius r_c	1.5 m
	Dryden turbulence intensity	σ_w=3.0 m/s
Control Algorithm	Prediction/Control horizon N_p/N_c	20 / 5
	SAC learning rate / Batch size	3×10^{-4} / 256

5.2 Convergence Analysis of the SAC Autonomous Decision Agent

During the training phase of rendezvous decision-making, the proposed SAC algorithm was compared against the Deep Deterministic Policy Gradient (DDPG) algorithm. Due to the strong nonlinearity of the wake vortex flow field, the DDPG agent frequently triggered collision penalties by inadvertently entering the intense downwash region during exploration, causing it to fall into a local suboptimal policy at around 1500 episodes and fail to find a smooth entry corridor. Conversely, the **SAC algorithm, relying on its stochastic exploration capability driven by the maximum entropy mechanism**, stably converged to the global optimal reward domain after approximately 1800 episodes of online interaction. It successfully planned an optimal approach trajectory bypassing the core vortex region from the lower flank.

5.3 Verification and Comparison of High-Precision Trajectory Tracking in the Docking Phase

To highlight the superiority of the proposed control architecture, the **Fuzzy PID** proposed in the

original literature and the widely applied **Nonlinear Dynamic Inversion (NDI)**[7] were selected as baseline comparison algorithms for verification in identical scenarios.

5.3.1 Dynamic Response of Spatial Relative Pose Errors

As the system entered the final 10 m high-precision capture phase, the hose-drogue exhibited a nonlinear "whipping" high-frequency oscillation with an amplitude of approximately ± 0.8 m due to the tangential wind field of the wake vortex.

At this stage, lacking the capability for predictive compensation of time-varying disturbances, and with expert rules struggling to cover the full-envelope aerodynamic nonlinearities, the **Fuzzy PID controller** experienced severe phase lag while tracking the drogue. Its maximum overshoots in the Y and Z axes reached 1.45 m and 1.38 m, respectively, causing the refueling probe to repeatedly scrape the edge of the drogue without achieving a lock-on. Facing unmodeled aerodynamic interference caused by the wake vortex, the **NDI controller** triggered high-frequency attitude oscillations due to inversion model mismatch, with the maximum overshoot still reaching 1.25 m.

In contrast, the **SAC-AMPC controller** proposed in this paper smoothly corrected the UAV's heading because the first-order incremental observer continuously injected feedforward anti-disturbance compensation into the system, while the MPC anticipated the kinematic trends of the drogue over a prediction horizon of 20 steps (N_p=20). The maximum transient overshoot throughout the docking process was strictly bounded within 0.55 m. Around 21.6 s, the three-axis relative errors were successfully reduced and maintained within the lock-on tolerance of ≤ 0.3 m.[8]

5.3.2 Handling Capability of Physical Hard Constraints

At t=12 s in the simulation, a transient crosswind was artificially superimposed. The integral term of the Fuzzy PID algorithm accumulated sharply (windup), leading to an aileron deflection command exceeding 35°. In a real physical system, this would directly trigger actuator saturation ($\pm 25^{\circ}$), causing the control loop to completely fail. However, the AMPC handled the constraints in Equation (8) strictly as boundary conditions for the QP optimization, ensuring all deflection commands remained safely within the flight envelope and eliminating instability caused by saturation.

5.4 Quantitative Metrics and Monte Carlo Statistical Analysis

A total of 100 Monte Carlo shooting simulations with random initial disturbance values were conducted. A rigorous quantitative statistical analysis of the Root Mean Square Error (RMSE), energy consumption (the time integral of squared actuator control effort Ju), and docking success rate was performed for each algorithm, as presented in Tables 2 and 3.

Table 2 Comparison of Trajectory Tracking RMSE and Control Energy Consumption

Control Algorithm	X-axis RMSE (m)	Y-axis RMSE (m)	Z-axis RMSE (m)	Energy J_u ($\times 10^3$)	Stable Docking Time (s)
Conventional PID	1.45	1.76	1.95	16.2	> 40 (Diverged)
Fuzzy PID (Baseline)	0.92	1.08	1.35	12.5	38.5
NDI Control	0.75	0.88	1.05	14.8	32.4
SAC-AMPC (Proposed)	**0.35**	**0.42**	**0.48**	**8.9**	**21.6**

Table 3 Dynamic Performance and 100 Monte Carlo Runs Statistics

Performance Metric	Fuzzy PID	NDI Control	SAC-AMPC (Proposed)
Max Transient Overshoot (Y/Z) (m)	1.45	1.25	**0.55**
Steady-state Pitch Variance (°)	2.1	1.6	**0.6**
Actuator Saturation Rate	12%	9%	**0%**
Precise Docking Success Rate	76.0%	82.0%	**98.0%**

In-Depth Data Analysis:

1) **Quantum Leap in Accuracy**: As shown in Table 2, compared with the Fuzzy PID control from the literature, the trajectory tracking RMSE of the SAC-AMPC architecture along the X, Y, and Z axes was significantly **reduced by 61.9%, 61.1%, and 64.4%**, respectively, demonstrating exceptional resilience to wind fields.

2) **Tremendous Optimization in Actuator Efficiency**: The rolling optimization mechanism of the MPC effectively filters out high-frequency clutter by penalizing the control increment Δu in the cost function. Compared to Fuzzy PID, the system control energy J_u **decreased by 28.8%**. This not only substantially reduces fatigue wear on physical execution elements such as servos but also endows the flight platform with higher cruising economy.

3) **Extreme Reliability Under Harsh Conditions**: Table 3 visually indicates that under strong wake vortex interference, conventional controls are highly prone to divergence due to actuator saturation (12% trigger rate), managing a docking success rate of only 76.0%. The proposed method completely eradicated the saturation-induced loss of control (0% trigger rate), achieving a **98.0% docking success rate** across a hundred extreme tests, fully satisfying the stringent rigor and reliability industrial standards required for the autonomous control of modern UAVs.

6. Conclusions

Targeting the extreme aerodynamic disturbances of wake vortices, the high-frequency motion of the drogue, and complex physical constraint problems faced during the UAV aerial refueling docking phase, this paper overthrows traditional paradigms relying on fixed patterns and linear control laws. An innovative cascaded framework based on SAC reinforcement learning and Adaptive Model Predictive Control (AMPC) is proposed and verified. The following conclusions are drawn:

1) **Implementing global rendezvous decision-making relying on the SAC algorithm equipped with a "maximum entropy" mechanism** effectively overcomes the limitations of traditional state-machine logic in dealing with complex flow fields. The agent is capable of online exploration within an unknown strong downwash environment, planning an optimal collision-avoidance entry trajectory that circumvents the vortex core.

2) **The constructed AMPC predictive controller, combined with the incremental disturbance observer**, achieves coordinated feedforward-feedback control against the "whipping" high-frequency dynamics of the drogue. By enforcing hard constraint limits on the execution elements at the QP solver level, it thoroughly eliminates the integrator windup and extreme deflection problems frequently occurring in traditional Fuzzy PID, compressing the maximum tracking overshoot into an extremely low safety margin (under 0.55 m).

3) **Rigorous dynamic simulations indicate** that this cascaded architecture stably controls the three-axis docking tracking RMSE of the receiver UAV under the combined action of turbulence and wake vortices at a 0.48 m level. The success rate across 100 shooting tests reaches a staggering 98.0%, accompanied by a 28.8% drop in control energy consumption. This research significantly improves the wind resistance and reliability of fully autonomous aerial refueling systems for UAVs

in strongly disturbed environments, providing a solid theoretical foundation and a highly scalable technical reference for intelligent cluster operations of next-generation aircraft.

References

[1] LIN X K, LIANG X L, REN B X, et al. Aerial refueling trajectory tracking control for UAV based on fuzzy PID [J]. Journal of Ordnance Equipment Engineering, 2022, 43(10): 18-26.

[2] DUAN H B, LI J, YIN S. An autonomous rendezvous and docking framework for UAV aerial refueling using deep reinforcement learning [J]. IEEE Transactions on Aerospace and Electronic Systems, 2021, 57(3): 1465-1478.

[3] ZHANG K, WANG Z, TENG G. Active disturbance rejection control for autonomous aerial refueling of UAVs under varying wake vortex [J]. Aerospace Science and Technology, 2023, 134: 108152.

[4] LI L, CHEN Y, MENG Y. Adaptive Model Predictive Control for constrained quadrotor tracking under unknown aerodynamic disturbances [J]. Automatica, 2022, 140: 110221.

[5] HAARNOJA T, ZHOU A, ABBEEL P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor [C]// International Conference on Machine Learning (ICML). PMLR, 2018: 1861-1870.

[6] BURNHAM D C, HALLOCK J N. Chicago monostatic acoustic vortex sensing system, volume IV: Wake vortex decay [R]. FAA, 1982.

[7] SUN J, CHEN Z, SHEN H. Robust nonlinear dynamic inversion control for UAV close-formation flight in turbulence [J]. Acta Astronautica, 2020, 175: 399-411.

[8] LU Y P, YANG C X, LIU Y Y. A survey of modeling and control technologies for aerial refueling system [J]. Acta Aeronautica et Astronautica Sinica, 2014, 35(09): 2375-2389.