Journal of Zhejiang University-SCIENCE A 2026 27(2):128-141
www.jzus.zju.edu.cn; www.springer.com/journal/11582
E-mail: jzus_a@zju.edu.cn

Research Article

https://doi.org/10.1631/jzus.A2400567

Efficient sensorimotor cues for training a glider to soar autonomously

Siyuan ZHENG[1], Jiachi ZHAO[1✉], Lifang ZENG[1✉], Zhouhong WANG[1], Jun LI[1,2]

[1]*School of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China*
[2]*Huanjiang Laboratory, Shaoxing 311800, China*

Abstract: Migratory birds depend on the perception of atmospheric updraft for long-distance flight. To realize more efficient autonomous soaring in an unpowered glider, different strategies for using potential sensorimotor cues to achieve autonomous soaring efficiency were compared and optimized. A simulation framework of autonomous soaring for an unpowered glider was developed based on a reinforcement learning algorithm. The framework was composed of three models: an updraft environment model, the glider's dynamics and control model, and a reinforcement learning agent, which learns to harvest more energy in flight. Based on the simulation, effects of different combinations of 12 potential sensorimotor cues on soaring efficiency were studied. Firstly, the absence of one particular sensorimotor cue and the use of only a single valid cue in autonomous soaring were analyzed. The results showed that the vertical airflow velocity gradient (a_w) and the wing-tip updraft velocity difference (τ) have advantages over the other cues. Secondly, strategies combining a_w or τ with other cues were analyzed to achieve more effective autonomous soaring, and seven potentially effective combinations of sensorimotor cues were identified. The final results showed that, among the tested combinations, the combination of vertical airflow velocity (V_w) and τ, enables the most efficient autonomous soaring. This study identified a highly effective sensorimotor cue strategy to guide an intelligent glider to achieve long-distance autonomous soaring flight.

Key words: Autonomous soaring; Glider; Reinforcement learning; Twin delayed deep deterministic policy gradient (TD3); Sensorimotor cues

1 Introduction

Migratory birds often use atmospheric updrafts to obtain energy for long-distance flights (Huey and Deutsch, 2016; Reddy et al., 2016; Kahn, 2017). Updraft-based flight is known as soaring (Liu et al., 2021). Inspired by birds, there have been many studies of the flight mechanism of soaring. Biomimetics of soaring could be applied to aircraft to help achieve long-endurance flight with the consumption of extremely low amounts of energy. For unmanned aerial vehicles (UAVs), developing a flight strategy based on atmospheric updrafts is the first step towards autonomous soaring.

Like birds, bionic UAVs need to have three capabilities to realize autonomous soaring: atmosphere perception, autonomous flight control, and trajectory planning. In terms of perception, UAVs should be able to estimate the surrounding airflow and current flight status according to the information obtained by sensors. Then, the flight control system makes corresponding flight decisions. Finally, the UAVs should track the planning path.

Atmospheric perception is the basis of the autonomous soaring of UAVs. Efficient algorithms are essential for accurate estimation of surrounding airflow characteristics. To achieve high-precision airflow estimation, diversified airflow sensing and estimation methods were proposed. To achieve real-time and accurate wind field estimation, Kahn (2017) proposed a localized estimation method for atmospheric heat flow, which uses an energy variation meter to obtain detailed updraft data and locate the center of the heat flow. Langelaan

✉ Jiachi ZHAO, jiachizhao@outlook.com
Lifang ZENG, lifang_zeng@zju.edu.cn
Siyuan ZHENG, https://orcid.org/0009-0001-4796-6094
Jiachi ZHAO, https://orcid.org/0000-0003-2404-3104
Lifang ZENG, https://orcid.org/0000-0001-5378-1208
Zhouhong WANG, https://orcid.org/0009-0001-5775-5418

Received Dec. 9, 2024; Revision accepted July 23, 2025;
Crosschecked Dec. 19, 2025; Online first Jan. 22, 2026

et al. (2011) investigated several methods for wind field estimation on small UAVs, which calculated the wind using global positioning system (GPS) and airspeed measurements, and tested these methods using a three-degree-of-freedom model in simulation. A real-time wind field map drawing technique was presented by Lawrance and Sukkarieh (2011) and used on an unpowered UAV flight. The precision and quality of the wind field map were confirmed by three simulated trajectories. Chan et al. (2011) laid out three methods for estimating the local wind speeds around fixed-wing UAVs in real-time, which were calculated by the navigation equation based on the data of ground speed and ascent rate from GPS, using the extended Kalman filter and the unscented Kalman filter. The outcomes confirmed the validity of the suggested techniques. Rhudy et al. (2013) developed a wind field estimation method based on pitot tubes and airflow sensors and rigorously evaluated the method in flight tests, proving its effectiveness. Moore et al. (2012) put forth a pure vision-based optimization method for wind field estimation that determines the characteristics of the wind field only through continuous measurements of the aircraft's heading and ground trajectory. Both simulations and flight tests validated the accuracy and robustness of the method, even under uncontrolled environmental conditions.

Traditional research methods for autonomous UAV flight decision-making rely on locating the center of the updraft or using simplified wind field environments and UAV models. Allen (2005) proposed a flight method to track the center of heat flow, in which he applied a cyclic strategy to obtain energy from the center of the heat flow. Simulation results showed that a UAV could greatly improve its endurance by using the updraft in the atmosphere. Depenbusch et al. (2018a, 2018b) proposed an algorithm that can be used for autonomous soaring of aircraft. The algorithm framework includes the mapping of heat flow and the decision-making to explore and use the wind. The feasibility of this algorithm was validated by actual flight tests. Edwards and Silverberg (2010) developed a new method to locate and keep a glider in an upwelling heat flow. They participated in a drone soaring competition, in which they compared the effectiveness of artificial and autonomous soaring.

In recent years, more and more researchers have used intelligent algorithms in autonomous static soaring strategies for UAVs. Autonomous static soaring relies on the perception and estimation of airflow to use the updraft to ascend and harvest energy. To gain more energy in soaring, drones need to choose between exploring unknown updrafts and using currently perceived updrafts, which is similar to the problem faced by agents in reinforcement learning methods on balancing exploration and utilization in uncertain environments. In the field of reinforcement learning, methods of balancing exploration and utilization have been widely and deeply studied. Aviation researchers have gradually begun to use reinforcement learning methods to solve the problem in autonomous soaring.

Reddy et al. (2016, 2018) investigated the soaring problem in an ascending turbulent airflow environment by combining digital simulation technologies with reinforcement learning techniques. Through simulation, they acquired sensory cues and practical soaring tactics that can efficiently guide soaring in a turbulent environment, realizing autonomous UAV navigation. Subsequently, they conducted field flight tests with great success and proposed an autonomous soaring technique that works in tumultuous environments. Notter et al. (2019) presented a model-free approach based on the Markov decision process (MDP) for flying in a perpendicular plane heat flow. The results of simulations showed that the trained agent's control may reasonably behave in a random environment and imitate the best behavior. A unique hierarchical reinforcement learning technique was described in their follow-up study to offer a control strategy for the competitive task of remotely piloting a glider (Notter et al., 2021). Furthermore, they suggested representing the reinforcement learning approach for the use of updrafts using a deep learning network with a loop control strategy structure (Notter et al., 2022). Woodbury et al. (2014) used reinforcement learning methods to control the tilt angle, navigate the glider to an updraft with a known position, and perform a hovering flight to gain energy. Chung et al. (2015) studied temporal difference reinforcement learning with adaptive and directional exploration under limited resources and proposed the enhanced Gaussian process–state–action–reward–state–action (eGP-SARSA) algorithm.

To obtain more energy from the updraft, optimal energy harvesting strategies were studied. In the MacCready method (MacCready, 1958), the optimal soaring speed is expressed as a function of the glider's

sinking speed and the thermal's ascent speed to maximize the average speed of a UAV. Walton et al. (2018) studied the optimal trajectory of a glider to obtain altitude from a thermal updraft. They compared different circling modes to find those that allow the glider to obtain more energy. Edwards et al. (2016) studied energy efficiency in different spiral trajectories and thermal updrafts.

As mentioned above, static soaring can effectively extend the endurance of UAVs. Reinforcement learning algorithms have been successfully used for autonomous soaring, and good results have been achieved. However, for the reinforcement learning algorithm, the output action depends on the current state, which is the flight characteristics and external airflow environment characteristics that the glider agent can observe in soaring. The impact of these characteristics on the soaring performance of gliders is not yet clear. To obtain a more effective and concise soaring strategy, in this study, we investigated the influence of sensorimotor cues on soaring efficiency. The main purpose was to determine which cues among the glider's flight characteristics and external environment cues are the most effective for guiding the glider to harvest more energy in a specific environment. Meanwhile, the selected cues should minimize the need for on-board sensors, which means the number of sensorimotor cues used should be as low as possible.

Firstly, a simulation framework of autonomous soaring for an unpowered glider was developed based on a twin delayed deep deterministic policy gradient (TD3) reinforcement learning algorithm. Then, the effects of nine flight characteristics and three updraft characteristics on the soaring efficiency of the glider were analyzed and compared. Finally, the sensorimotor cues that were the most efficient for glider soaring in the specific environment were identified.

2 Methods and modelling

The basic framework is illustrated in Fig. 1. There are three main modules: the updraft environment model, the glider model, and the reinforcement learning agent. The updraft environment is uniformly distributed. Additionally, a detailed six-degree-of-freedom (6-DOF) glider dynamics model is developed based on the physical characteristics of the glider. Data from the interaction between the updraft environment and the glider model form the basic clues for the reinforcement learning training. Twelve flight state parameters are considered here to determine the most useful sensorimotor cues for autonomous soaring. The TD3 algorithm is used to train the glider agent.

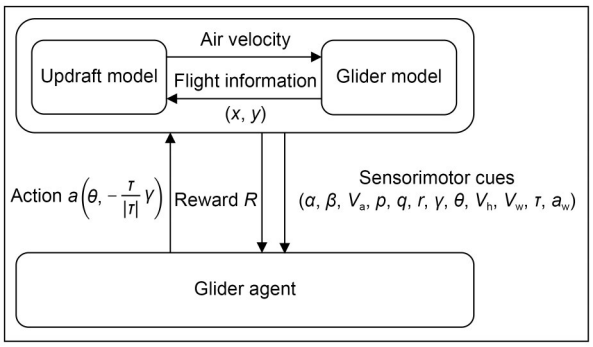

Fig. 1 Autonomous soaring research framework using reinforcement learning. The parameters are explained in the text

2.1 Glider model

The glider is a conventional UAV with high aspect ratio wings. The inner wing is straight, while the outer sections are trapezoidal with a 6° sweep angle at the leading edge. A 3D model of the glider is illustrated in Fig. 2, and the main geometric parameters are listed in Table 1.

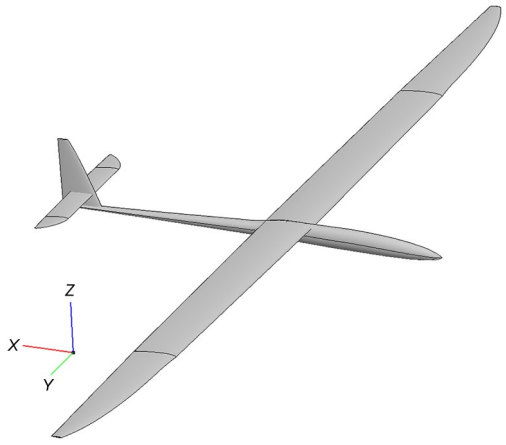

Fig. 2 3D geometric model of the glider

Based on the 3D geometric shape, the aerodynamic parameters were estimated using the open-source software OpenVSP. Aerodynamic forces and moments, calculated at various angles of attack (α), were used to construct the aerodynamic equations.

Table 1 Geometric parameters of the glider model

Parameter	Description
Take-off weight	5.5 kg
Fuselage length	1.9 m
Aspect ratio of wing	26.56
Wing area	0.9225 m^2
Wingspan/airfoil	4.95 m/MH-32
Horizontal tail span/airfoil	0.845 m/NACA0010
Vertical tail span/airfoil	0.33 m/NACA0010

In each simulation step, according to the glider's current flight state, the corresponding aerodynamic coefficients are calculated from the aerodynamic equations. Then, the aerodynamic and inertial forces and moments are input into the dynamic equations to update the glider's state. This process connects the aerodynamic characteristics with flight dynamics. A 6-DOF dynamic equation of the glider was built, formulated as:

$$\begin{cases} F_{x_b}=m(\dot{u}+qw-rv), \\ F_{y_b}=m(\dot{v}+ru-pw), \\ F_{z_b}=m(\dot{w}+pv-qu), \end{cases} \tag{1}$$

$$\begin{cases} M_{x_b}=I_{xx}\dot{p}-I_{zx}(\dot{r}+pq)-(I_{yy}-I_{zz})qr, \\ M_{y_b}=I_{yy}\dot{q}-I_{zx}(r^2-p^2)-(I_{zz}-I_{xx})rp, \\ M_{z_b}=I_{zz}\dot{r}-I_{zx}(\dot{p}-qr)-(I_{xx}-I_{yy})pq, \end{cases} \tag{2}$$

where F_{x_b}, F_{y_b}, and F_{z_b} are the force components in the body frame, m is the mass of the glider, M_{x_b}, M_{y_b}, and M_{z_b} are the aerodynamic moments, I_{xx}, I_{yy}, I_{zz}, and I_{zx} are the moments of inertia, u, v, and w are the velocity components, and p, q, and r are the angular velocity components.

This closed-loop process ensures a tight coupling between the aerodynamic characteristics and flight dynamics throughout the simulation.

Based on the aerodynamic parameters in Table 2, a 6-DOF glider dynamics model was developed. Numerous autonomous flight simulations have defined the flight parameter ranges for glider models (Wharington and Herszberg, 1998; Walton et al., 2018; Powers et al., 2020). Based on these studies, the following bounds were applied to the state values:

$$\begin{aligned} &-60°\leqslant\gamma\leqslant 60°,\ -10°\leqslant\theta\leqslant 10°, \\ &|h|\geqslant 20,\ 5\leqslant V_a\leqslant 30,\ \alpha\leqslant 14°, \end{aligned} \tag{3}$$

where γ represents the glider's roll angle, θ denotes the pitch angle, h is the flight altitude (m), V_a is the airspeed (m/s), and α is the angle of attack (AOA). For different gliders, these boundaries should be re-evaluated according to each aircraft's characteristics.

Table 2 Aerodynamic characteristics of the glider model

Parameter	Value
Lift coefficient at zero AOA $\left(C_{L0}=\frac{2L}{\rho V^2 S}\right)$	0.1284
Lift coefficient due to AOA ($C_{L\alpha}$)	0.1132
Zero-lift drag coefficient $\left(C_{D0}=\frac{2D}{\rho V^2 S}\right)$	0.0230
Pitch moment coefficient at zero AOA $\left(C_{M0}=\frac{2M}{\rho V^2 \bar{c}}\right)$	−0.0421
Pitch moment coefficient due to AOA ($C_{M\alpha}$)	−0.0226

L: lift force; D: drag force; M: pitching moment; ρ: air density; V: airspeed; S: wing area; $\bar{c}$: mean aerodynamic chord

In this study, a proportional–integral (PI) feedforward controller was implemented to regulate θ and γ of the glider. The PI feedforward controller acts as a low-level attitude controller, tracking the target angles generated by the reinforcement learning agent and generating actuator commands to minimize the error between the target and actual angles. The control structure is illustrated in Fig. 3. The control equation is formulated as follows:

$$\tilde{u}(t)=K_p e(t)+K_i\int_0^t e(t)\mathrm{d}t+K_f\omega, \tag{4}$$

where $\tilde{u}(t)$ is the output signal, and K_p, K_i, and K_f are the proportional, integral, and feedforward gains, respectively. $e(t)$ is the error between the target angle and the actual angle at time t. The variable ω represents the feedback angular velocity used in the feedforward loop (corresponding to the pitch rate q or roll rate p in the body frame). In the control block diagram (Fig. 3), this rate feedback is denoted as the time derivative of the angle ($\dot{\theta}$, $\dot{\gamma}$).

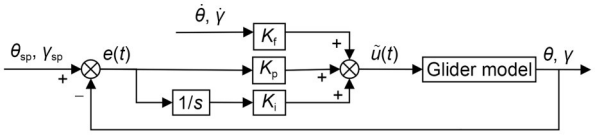

Fig. 3 6-DOF glider dynamics model and PI feedforward controller

Regarding the notations in Fig. 3, the subscript 'sp' indicates the setpoint angles generated by the reinforcement learning agent. The signs '+' and '−' at the summing junction represent the calculation of the error $e(t)$ by subtracting the actual angle from the target angle. Furthermore, the symbol s in the block diagram denotes the Laplace operator for integration.

The PI feedforward controller gains (K_p, K_i, and K_f) directly influence the glider's response and tracking accuracy during changes in roll and pitch angles. Higher K_p values improve responsiveness, K_i helps eliminate steady-state error, and K_f enhances the tracking of target commands. Proper tuning ensures stable and precise attitude control. The final gains of PI feedforward controller are listed in Table 3.

Table 3 Final gains for the PI feedforward controller

Loop	K_p	K_i	K_f
θ	2.00	1.50	0.30
γ	0.80	0.60	0.15

2.2 Environment model

The environment model is a simplified round updraft model with the following mathematical expression:

$$V_w(x, y) = w_0 \cdot e^{-\frac{(x-x_c)^2+(y-y_c)^2}{2500}}, \quad (5)$$

where V_w denotes the updraft velocity (unit: m/s) and w_0 indicates the maximum updraft velocity. The coordinates x and y represent the glider's position at the ground reference coordinate system. The center of the updraft (x_c, y_c) is positioned at (0, 0), where the vertical air velocity reaches its peak value, w_0.

Fig. 4 illustrates the distribution of the round updraft when w_0=8 m/s. The round updraft follows a typical Gaussian distribution, with the airflow velocity exhibiting clear symmetry around the wind field center. In this updraft scenario, the glider must quickly navigate from its initial position to the airflow center and maintain an optimal attitude angle to maximize energy acquisition.

In this study, the model was trained and tested in a simplified circular updraft environment, which is one of the fundamental wind fields in the natural environment. We have also tested the glider agent in other complex wind fields, such as the Gedeon thermal model and Dryden wind turbulence model. The results show that

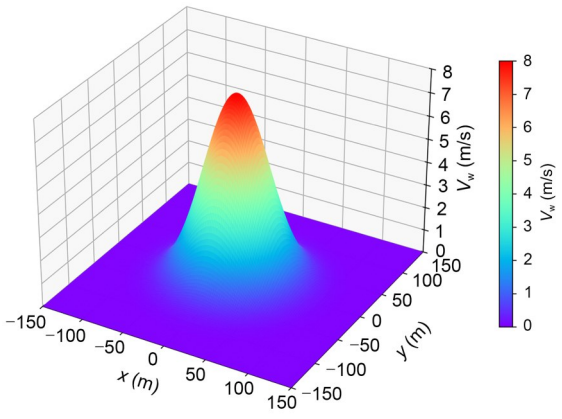

Fig. 4 Velocity distribution of the round updraft model. References to color refer to the online version of this figure

the agent trained only in the simplified circular updraft demonstrates strong generalization ability and robustness in various complex airflow environments. Further details can be found in reference (Zhao et al., 2023).

2.3 Reinforcement learning algorithm

Reinforcement learning is a subset of machine learning, driven mainly by the interaction between an agent and the environment, allowing the agent to be continuously updated and refined. The objective is to learn behaviors that maximize the reward. At each time step k, the agent acquires the state o and selects an action a based on its policy π. Following each action, the agent receives the reward $\tilde{r}$ and transitions to a new state o'. The goal of reinforcement learning is to maximize the cumulative discounted future reward (R_k) for each state o.

$$R_k = \sum_{i=k}^{k_{max}} \mu^{i-k} \tilde{r}(o_i, a_i), \quad (6)$$

where μ represents the discount factor (with $0 \leqslant \mu \leqslant 1$) and k_{max} denotes the maximum number of time steps in an episode.

Most studies on autonomous soaring have used discrete reinforcement learning algorithms to control agents (Woodbury et al., 2014; Chung et al., 2015; Reddy et al., 2016), where the action space is composed of predefined actions. In the case of glider autonomous flight, many discrete reinforcement learning control methods adjust the current attitude angle with a specific value. Alternatively, they maintain the attitude angle unchanged, which limits the glider's

flight performance. In this study, a continuous reinforcement learning algorithm, TD3, was implemented to set the target attitude angle within a specified range to enhance the flight capabilities.

The TD3 algorithm is a deep reinforcement learning algorithm based on the actor–critic framework. It considers the interaction between function approximation errors in both policy and value updates. TD3 is an improvement over the deep deterministic policy gradient (DDPG) algorithm, which aims to reduce the tendency of DDPG to overestimate in evaluating the current strategy. TD3 comprises six neural networks: two critic networks, Q_{ϑ_1}, Q_{ϑ_2}, two target critic networks, $Q_{\vartheta'_1}$, $Q_{\vartheta'_2}$, the actor-network $\pi_{\varnothing}$, and the target actor-network $\pi_{\varnothing'}$.

To start the train, critical networks Q_{ϑ_1}, Q_{ϑ_2}, and actor-network $\pi_{\varnothing}$ are initialized with random parameters ϑ_1, ϑ_2, and $\varnothing$, while the target networks are initialized with $\vartheta'_1 \leftarrow \vartheta_1$, $\vartheta'_2 \leftarrow \vartheta_2$, and $\varnothing' \leftarrow \varnothing$. Simultaneously, TD3 initializes a replay buffer B to store the past experiences.

At each time step k, the action a is first executed by the agent, which causes the system to enter a new state o'. The agent then receives the reward $\tilde{r}$ from the environment at that time step, and the state transition tuple (o, a, $\tilde{r}$, o') is stored in the replay buffer B. A mini-batch of N transition tuples is randomly sampled from B. To address value overestimation, target policy smoothing is applied to obtain the target action:

$$\tilde{a} = \pi_{\varnothing'}(o) + \epsilon, \tag{7}$$

$$y_{\text{target}} = \tilde{r} + \mu \min_{i=1,2} Q_{\vartheta'_i}(o',\ \tilde{a}), \tag{8}$$

where $\tilde{a}$ represents the smoothed target action, ϵ is a clipped noise used for regularization, and μ is a discounted factor determining the priority of short-term rewards. Note that y_{target} represents the target Q-value for the update. Then, parameters ϑ_1 and ϑ_2 of critic networks Q_{ϑ_1} and Q_{ϑ_2} are updated by minimizing the following loss function:

$$J(\vartheta_i) = \frac{1}{N}\sum_{j=1}^{N}\left(y_{\text{target},j} - Q_{\vartheta_i}(o_j,\ a_j)\right)^2, \tag{9}$$

where N is the denotes the mini-batch size.

After the time instance d, the weight $\varnothing$ of the actor networks is updated according to the deterministic policy gradient:

$$\nabla_{\varnothing} J(\varnothing) = \frac{1}{N}\sum \nabla_a Q_{\vartheta_1}(o,a)|_{a=\pi_{\varnothing}(o)} \nabla_{\varnothing}\pi_{\varnothing}(o). \tag{10}$$

The target networks are updated as follows:

$$\begin{cases} \vartheta'_i \leftarrow \tau_{\text{r}}\vartheta_i + (1-\tau_{\text{r}})\vartheta'_i, \\ \varnothing' \leftarrow \tau_{\text{r}}\varnothing + (1-\tau_{\text{r}})\varnothing', \end{cases} \tag{11}$$

where τ_{r} represents the update rate.

2.3.1 Action space

Our method enables continuous control of the glider's θ and γ, with the variation range for θ defined as $\theta \in [-10°,\ 10°]$ and for γ as $\gamma \in [-60°,\ 60°]$.

2.3.2 Reward function

The goal of autonomous soaring is to maximize the energy collected by the unpowered glider in the updraft. Therefore, the reward function is designed as the gain in the glider's total energy, the sum of gravitational potential energy and kinetic energy per unit mass, which is expressed as

$$R_t = g(h_t - h_{t-1}) + \frac{1}{2}(V_t^2 - V_{t-1}^2), \tag{12}$$

where g is gravitational acceleration, R_t is the reward at time t, and h_t and V_t represent glider altitude and airspeed at time t, respectively.

2.3.3 Break conditions

In line with the reward function, the break conditions are based on the glider's altitude and airspeed. The minimum altitude $h_{\min}$ is set as 20 m to prevent the glider from touching the ground. The maximum altitude is unrestricted to allow for more gravitational potential energy. The glider operates within a broad airspeed range, with a minimum airspeed $V_{\min}$ of 6 m/s and a maximum airspeed $V_{\max}$ of 30 m/s.

Additionally, a maximum episode time is imposed to avoid excessively long episodes. The flight time per episode is capped at 100 s, ensuring enough time for the glider to reach the center of the airflow.

The current training episode ends whenever any of the three break conditions above are reached during the glider's training.

2.3.4 Training method

In each training round, the initial altitude of the glider is set as 100 m and the w_0 as 8 m/s. The initial θ, γ, and yaw angle φ are all set as 0°, with the initial horizontal position fixed at (−50 m, −50 m) to simplify the training process. Additionally, all initial conditions for all the cases are kept the same.

The time step for the TD3 algorithm is 1 s, while the update interval for the critic network, target critic network, actor network, and target actor network is 2 s. The glider's initial airspeed is 8 m/s. Three different initial random number seeds are used to train the TD3 agents separately. After training, three distinct intelligent agents are obtained. The one that maximizes the glider's mass-independent energy gain (R) is selected for further analysis. Note that this strategy is trained for the specific glider model described above; for other UAVs, retraining or fine-tuning may be required to ensure optimal performance.

2.4 Sensorimotor cues

A sensorimotor cue is perception information acquired from airborne sensors, which can be used for autonomous soaring. In this study, we applied reinforcement learning to explore the flight strategy of autonomous soaring. Based on some of the information perceived by the glider, the reinforcement learning agent decides to take action with certain γ and θ. Appropriate perception information is crucial for autonomous soaring, as the glider's intelligent agent relies on this information to make informed decisions. To minimize the need for electronic sensory devices, the selection of perception information is crucial and should follow several principles. Firstly, the perceived information must be sufficient for the glider to make appropriate autonomous soaring decisions. Secondly, the perceived information should preferably be observable by the glider in real time. Considering both the flight characteristics of gliders and the external airflow properties of unmanned aerial vehicles, 12 candidates of sensorimotor cues were compared to explore the most efficient strategy for autonomous soaring.

The candidates included the angular velocities p, q, and r, vertical velocity V_h in the Earth's coordinate system, airspeed V_a, and aerodynamic angles: α, sideslip angle β, θ, and γ. For external airflow characteristics, referring to Reddy et al. (2016), the vertical updraft acceleration a_w, the updraft velocity V_w, and the wing-tip updraft velocity difference τ were also chosen as candidate sensorimotor cues. The details of these 12 investigated sensorimotor cues are summarized in Table 4. The variable τ is defined as:

$$\tau = V_{wl} - V_{wr}, \tag{13}$$

where V_{wl} and V_{wr} denote the difference in vertical updraft velocities at the left and right-wing tips, respectively.

Table 4 Twelve investigated sensorimotor cues

Sensorimotor cue	Symbol
Angular velocity	p, q, r
Vertical velocity	V_h
Airspeed	V_a
Aerodynamic angle	α, β, θ, γ
Vertical updraft acceleration	a_w
Updraft velocity	V_w
Wing-tip updraft velocity difference	τ

Among the 12 candidate sensorimotor cues, most, such as p, q, r, V_h, V_a, θ, and γ, can be directly measured by standard onboard sensors (e.g., inertial measurement units (IMUs), airspeed sensors, GPS/barometric altimeters). α and β can also be measured, though they typically require dedicated sensors such as multi-hole probes or differential pressure sensors, which may not be available on all UAV or glider platforms.

For certain cues related to external airflow, specifically, a_w, V_w, and τ, direct measurement with standard sensors is challenging. However, these cues can be estimated indirectly: V_w can be estimated by measuring the glider's vertical velocity and subtracting its sink rate (determined from the airspeed and polar curve); a_w can be obtained by differentiating the estimated V_w over time with appropriate filtering; τ can be inferred by analyzing differential wing loading or structural responses, rather than direct measurement at the wingtips.

In this study, these cues were used mainly for theoretical analysis and to explore the performance limits under ideal sensing conditions. For practical applications, it is important to ensure that the selected cues can be reliably and effectively acquired in real time, either directly or through robust estimation algorithms.

Additionally, since the values of a_w and τ were smaller than those of other sensorimotor cues, they

were scaled by a factor K (set to 100) to standardize all sensorimotor cues.

3 Results and discussion

To assess the influence of various sensorimotor cues on autonomous soaring efficiency and to identify the optimal combinations of cues for guiding the glider, a sensitivity analysis of missing a certain sensorimotor cue and using only a single sensorimotor cue is first provided. Then, several potential combinations of the 12 sensorimotor cues are used as the agent's state inputs for training the glider in the round updraft environment.

3.1 Results of missing a sensorimotor cue

To assess the impact of missing a sensorimotor cue on the efficiency of autonomous soaring, the invalid sensorimotor cue remains a constant in the TD3 agent, while the other 11 cues vary continuously (Fig. 5b). Tests with eight different initial positions were conducted, including the training position at (−50 m, −50 m) and seven other points arranged on a circle with a radius of $50\sqrt{2}$ m (Fig. 6), to comprehensively evaluate the robustness and adaptability of the trained agent in different spatial scenarios. For all test cases, the initial velocity of the glider was set along the positive x-axis to standardize the initial flight conditions and ensure a fair comparison among different test points. These tests were used to evaluate the effect of individual sensorimotor cues on the generalization of the trained glider agent. To select appropriate fixed values for evaluation within a reasonable range, at each initial position, the maximum, minimum, median, and average values of each sensorimotor cue obtained from the simulation test round shown in Fig. 7, along with the fixed value of 0, were recorded. These five values were used as the fixed parameters for each sensorimotor cue.

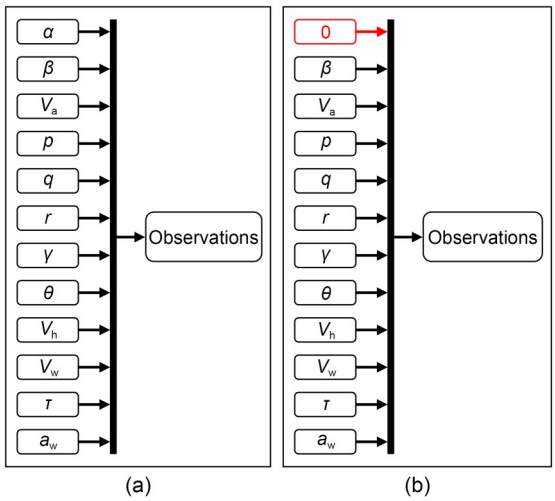

Fig. 5 Example of missing a sensorimotor cue (taking α as an example, the fixed value example is 0°): (a) all sensorimotor cues are variable; (b) the invalidated sensorimotor cue, α, is fixed at 0°

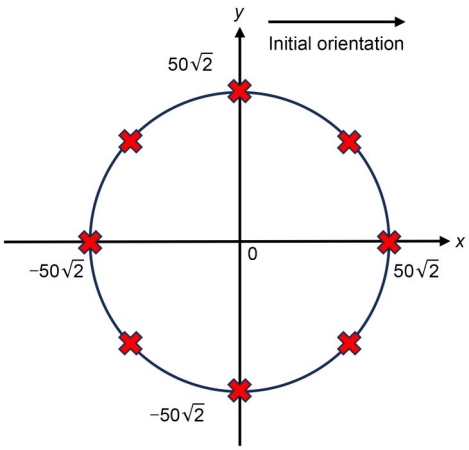

Fig. 6 Eight different initial positions of glider agents during different tests (unit: m)

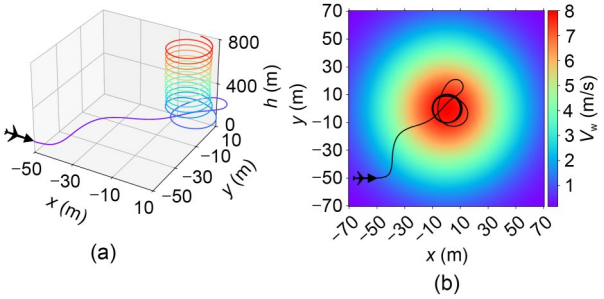

Fig. 7 Flight trajectory of the autonomous soaring glider: (a) side view of the glider's flight trajectory (velocities are identified by different colors); (b) top view of the glider's flight trajectory. References to color refer to the online version of this figure

At eight various initial positions, when different sensorimotor cues are missing, the distribution of average reward R_{avg} distribution is depicted in Fig. 8. Specifically, R_{avg} denotes the mean value of the energy gains calculated over the five fixed-value cases (maximum, minimum, median, average, and 0) defined above. As a comparison, the first column shows the R_{avg} value when all 12 sensorimotor cues can vary.

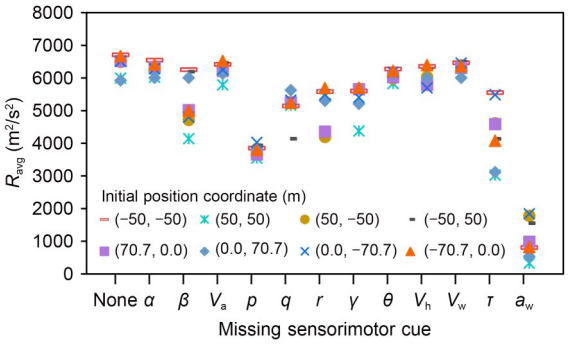

Fig. 8 Distribution of R_{avg} corresponding to missing a sensorimotor cue

The results reveal that the absence of a_w has a significant impact on autonomous soaring efficiency, making it nearly impossible for the trained agent to control the glider and accumulate energy. Furthermore, it is evident that the roll rate p and τ also have a noticeable effect on autonomous soaring performance.

3.2 Results from using only one valid sensorimotor cue

This section describes the effect on autonomous soaring efficiency of using an individual cue. In these cases, the tested sensorimotor cue was variable, while the remaining 11 cues were kept fixed (Fig. 9). The tests were conducted from the same eight initial positions depicted in Fig. 6. The case when all 12 sensorimotor cues are fixed serves as the baseline for comparison. At each initial position, the sensorimotor cue was set to the fixed value that yielded the highest reward R_{max} among the five fixed values mentioned above.

On this basis, the distribution of rewards R related to using a single sensorimotor cue is shown in Fig. 10. In the first column is the R value when all 12 sensorimotor cues remain fixed. The results show that a_w and τ have more effect than the other cues. These results agree with the findings of Reddy et al. (2016), which suggested that a_w and τ alone can guide gliders to achieve autonomous soaring. However, the other 10 cues struggle to enable the glider to achieve autonomous soaring.

A positive a_w means the vertical airflow speed is increasing in the direction of flight, while a negative a_w indicates a decrease in vertical airflow speed along the flight path. To reach the center of the updraft, the glider should fly in the direction where a_w is positive. Conversely, a negative a_w suggests the glider should

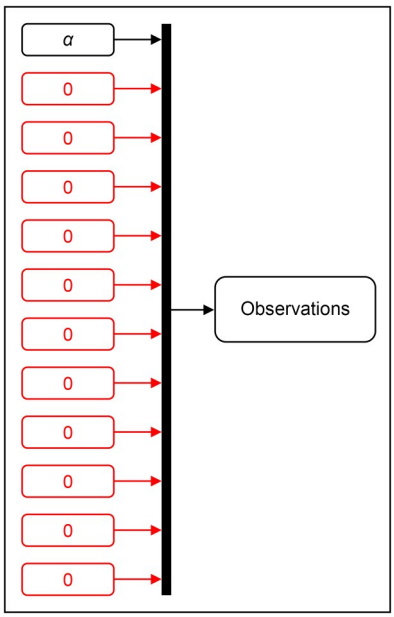

Fig. 9 Example of only one valid sensorimotor cue (taking α as an example, the fixed value example is 0)

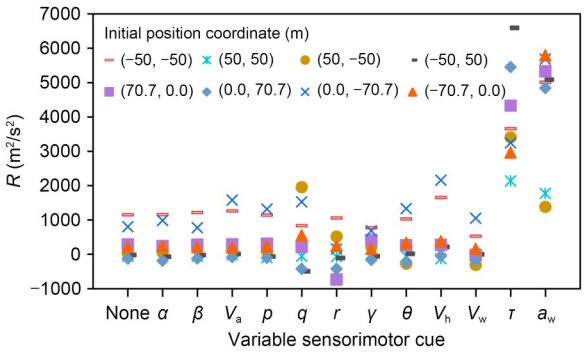

Fig. 10 Distribution of R (obtained with the fixed value yielding the maximum reward among the five candidates) corresponding to using only one valid sensorimotor cue

fly toward a region with lower vertical airflow velocity, requiring turning the direction to a higher updraft. However, the glider struggles to accurately determine the direction of the strongest updraft based solely on a_w. When a_w is negative (Fig. 11a), the glider needs to turn right, allowing it to reach the region with a stronger updraft and achieve a positive a_w. Nevertheless, relying only on a_w makes it difficult for the glider to turn in the correct direction. This highlights the need for additional sensorimotor cues to guide the flight, which can indicate the different intensities of vertical airflow velocities in different positions.

τ represents the difference in V_w between the left and right wings, which helps identify the potential

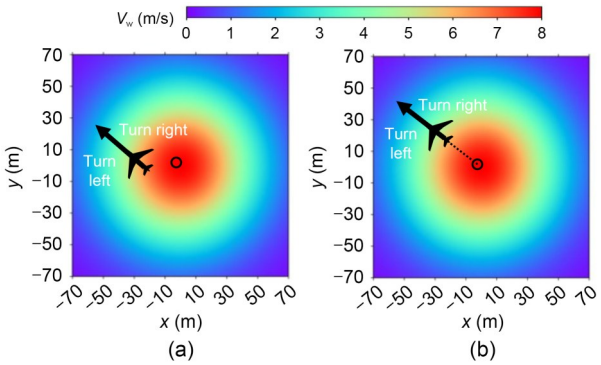

Fig. 11 Flight direction schematic: (a) negative a_w requires the glider to turn right with stronger updraft; (b) when τ= 0 m/s, it is difficult to predict V_w trends ahead based solely on τ. References to color refer to the online version of this figure

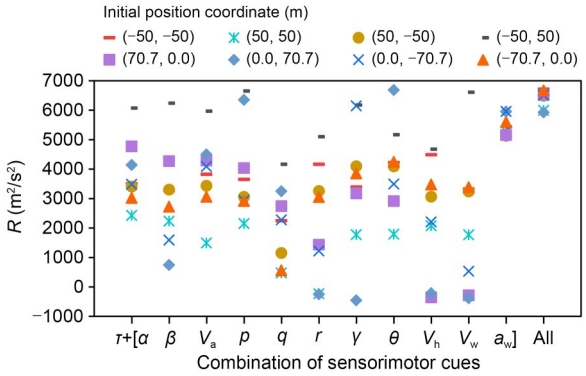

Fig. 12 Reward distribution for the combination of τ with different sensorimotor cues (the fixed value was selected to yield the maximum reward among the five candidates)

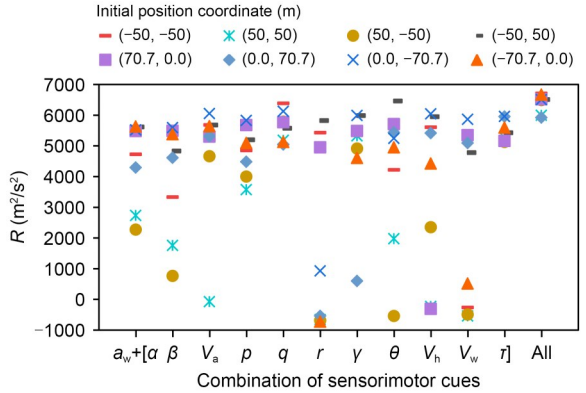

Fig. 13 Reward distribution for the combination of a_w with different sensorimotor cues (the fixed value was selected to yield the maximum reward among the five candidates)

direction of stronger vertical airflow, either on the left or right side of the glider's current position. In some special circumstances, for example when the glider's flight path is as depicted in Fig. 11b, where τ is equal to 0, it is difficult to rely solely on τ to determine the changing trend of V_w ahead. Therefore, other sensorimotor cues are required to assist in determining the turning direction.

3.3 Results of different sensorimotor cue combinations

3.3.1 Analysis of potential sensorimotor cue combinations

The results of Sections 3.1 and 3.2 show that both a_w and τ can independently guide the glider to successful autonomous soaring. However, relying on a single sensorimotor cue limits the glider's ability to make optimal decisions. To investigate more effective sensorimotor cue combinations for guiding autonomous soaring, the effect of combining a_w or τ with other cues on soaring performance was analyzed. The outcomes of simulations involving cues combining with τ and with a_w are presented in Figs. 12 and 13, respectively. Each case was tested with eight initial positions. Additionally, the reward under conditions where all 12 sensorimotor cues were used was compared, as shown in the last column of each figure. Clearly, the case using all 12 sensorimotor cues harvested the most energy.

Fig. 12 shows that, compared to the R value with all 12 cues used, the $\tau+p$ combination achieved a higher R value at the initial positions (−50 m, 50 m) and (0.0 m, 70.7 m), while $\tau+\theta$ showed better performance at the initial position (0 m, 70.7 m). Similarly, $\tau+V_w$ obtained a higher R value at (−50 m, 50 m), and $\tau+a_w$ performed better at (0.0 m, 70.7 m).

Furthermore, the data in Table 5 show that $\tau+a_w$, $\tau+\theta$, and $\tau+p$ gave the highest average reward values across the eight tested initial positions. Additionally, the combinations of $\tau+\theta$, $\tau+p$, and $\tau+V_w$ had the highest maximum reward values. Therefore, these four combinations were identified as potentially effective sensorimotor cue pairings for autonomous soaring and were investigated further.

Results in Fig. 13 show that, for combinations of a_w with other sensorimotor cues, the glider was unable to achieve a higher R than the one using all 12 sensorimotor cues. However, previous conclusions indicate that a_w possesses a certain ability to guide the glider for static soaring independently, and aside from the condition where all 12 cues vary, the $\tau+a_w$ combination

Table 5 Average and maximum rewards from eight initial evaluation positions

Sensorimotor cue combination	Average reward	Maximum reward	Sensorimotor cue combination	Average reward	Maximum reward
$\tau+\alpha$	3851	6083	$\tau+\gamma$	3523	6187
$\tau+\beta$	3044	6229	$\tau+\theta$	4075	6687
$\tau+V_a$	3833	5979	$\tau+V_h$	2427	4687
$\tau+p$	3973	6666	$\tau+V_w$	2270	6625
$\tau+q$	2104	4145	$\tau+a_w$	5403	5979
$\tau+r$	2221	5125	All	6414	6708

also yielded effective static soaring results. Therefore, to explore the effect of combining a_w with other cues on static soaring performance, the R values of different combinations were compared with that of $\tau+a_w$.

Through this comparison, the combinations of a_w+V_a, a_w+q, and $a_w+\gamma$ achieved higher rewards in five initial positions, which is a clear advantage over other combinations that achieved higher R values in at most three initial positions. Consequently, these three combinations were identified as potentially effective sensorimotor cue pairings for static soaring and were analyzed further.

3.3.2 Results of potential sensorimotor cue combinations on autonomous soaring

The previous study identified seven potential combinations of sensorimotor cues: $\tau+p$, $\tau+\theta$, $\tau+V_w$, $\tau+a_w$, a_w+V_a, a_w+q, and $a_w+\gamma$. Along with τ and a_w, which can independently guide the glider's soaring. Results from nine potentially effective sensorimotor cue combinations are compared here. Glider agents with these nine combinations were retrained individually to identify the best cue combinations.

The trends of the glider's θ and γ after training with two different initial positions are shown in Fig. 14. At the beginning of the flight, different initial positions require turns in opposite directions, so the glider issues opposite roll angle commands for the two initial positions. However, during the circling phase, the roll angles of the glider for both initial positions remain about the same, around −48°. These results indicate that increasing the diversity of initial training positions helps the glider agent to sample the updraft environment more evenly during training, thereby improving its learning performance.

Therefore, for each of the nine sensorimotor cue combinations, the glider agents were trained at three different initial positions (Fig. 15). The training positions

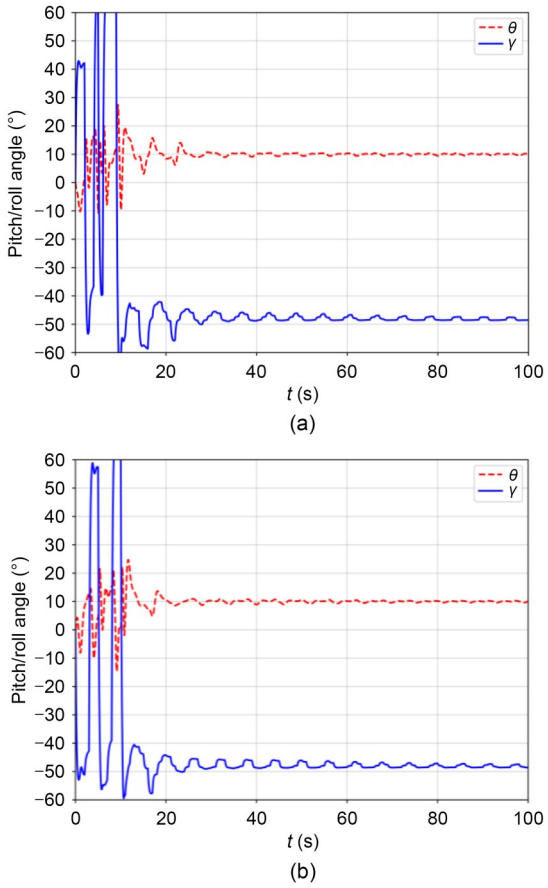

Fig. 14 Variation trends of pitch angle and roll angle of the glider at different initial positions: (a) (x_0, y_0)=(−50 m, −50 m); (b) (x_0, y_0)=(50 m, 50 m)

were located at coordinates (−50 m, −50 m), (50 m, −50 m), and (50 m, 50 m). When the initial position was at (−50 m, −50 m), the glider required only a minor turn to reach the center of the updraft. However, for initial positions at (50 m, −50 m) and (50 m, 50 m), the glider had to execute a large turn to reach the updraft center, making the agent's learning process more challenging. This training strategy not only increases the randomness beyond just varying the random

seed but also prevents overfitting to a specific initial condition.

After completing the training, 27 groups of agents were tested at eight different initial positions (evaluation points) (Fig. 6). The average reward achieved by the glider across these eight positions is presented in Fig. 16.

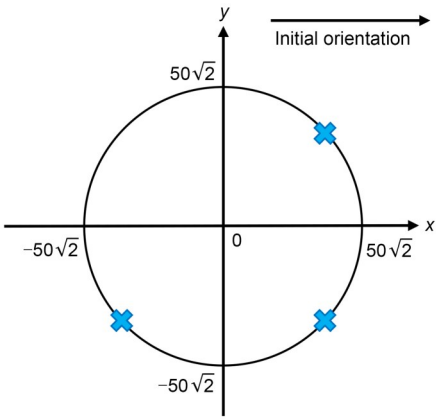

Fig. 15 Three distinct initial positions of glider agents during different tests (unit: m)

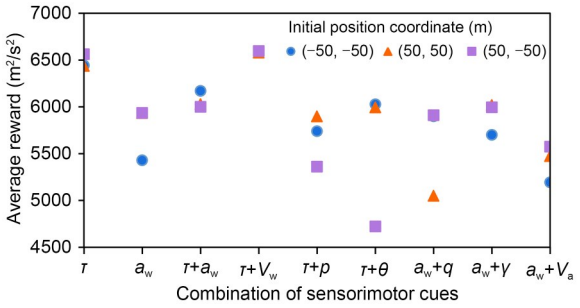

Fig. 16 Average rewards obtained by the agents in eight initial evaluation positions when trained at different initial positions

As shown in Fig. 16, the combination of $\tau+V_w$ achieved the highest average reward, indicating its superior ability to guide the glider. This aligns with the hypothesis that the glider requires the sensorimotor cues most capable of directing its turns and reflecting the intensity of vertical airflow velocity.

The combination of $\tau+a_w$ and τ alone also guided the glider effectively, but with slightly inferior results. Notably, all sensorimotor cues in the three most effective combinations are related to the updraft, whereas sensorimotor cues involving the glider's own flight parameters did not sufficiently support autonomous soaring.

A further comparison was conducted between $\tau+a_w$ and $\tau+V_w$. As shown in Fig. 17, although the glider guided by $\tau+a_w$ could identify the updraft center at different initial positions, its flight showed a significantly eccentric trajectory when circling around the center, resulting in relatively less energy gain. In contrast, the glider trajectory under $\tau+V_w$ guidance could circle around the origin at all initial positions, thereby collecting more energy. The training results show that in the round updraft model, the $\tau+V_w$ combination had a distinct advantage over the $\tau+a_w$ combination in guiding the glider to circle around the origin, enabling the glider to gain more energy. In conclusion, the sensorimotor cue combination of $\tau+V_w$ was the most effective for guiding the glider's autonomous soaring.

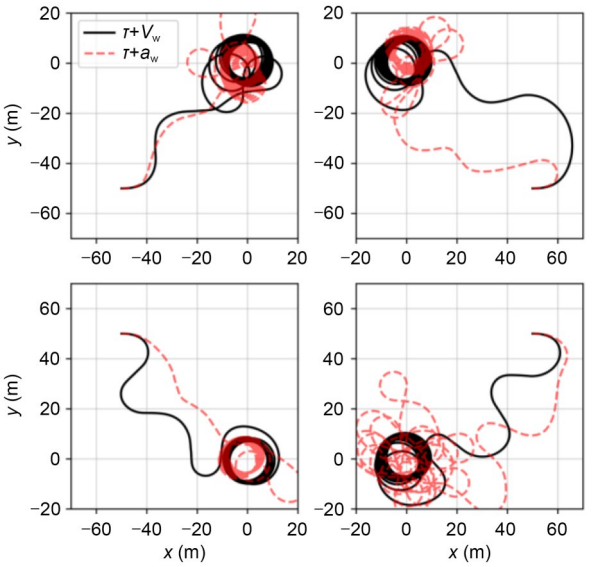

Fig. 17 Glider flight trajectories guided by different combinations of sensorimotor cues in round updrafts when trained at different initial positions

4 Conclusions

This study focused on the effects of sensorimotor cues on the autonomous soaring efficiency of a glider. Nine flight characteristics and three vertical airflow characteristics were identified as potential sensorimotor cues. A simulation framework for autonomous soaring of an unpowered glider was developed based on the TD3 reinforcement learning algorithm, in which 12 sensorimotor cues were used as the state inputs for the glider's intelligent agent. The agent was trained in

a circular updraft environment to maximize energy harvesting. Sensitivity analysis and combination analysis for different sensorimotor cues on autonomous soaring are discussed. The results are summarized as follows:

(1) Sensitivity analysis of missing one sensorimotor cue and using only a single sensorimotor cue was carried out. The results show that both a_w and τ could independently guide the glider to achieve autonomous soaring. However, the absence of a_w had the most significant negative impact on soaring efficiency. Relying on a single cue limits the agent's ability to make optimal decisions.

(2) The effect of combinations of a_w or τ with other cues on soaring performance was analyzed. The results show that the combination of V_w and τ consistently achieved the highest energy gain and the best soaring performance across diverse scenarios.

These findings provide valuable guidance for the selection of sensorimotor cues in intelligent soaring gliders. Appropriate sensorimotor cues will help gliders achieve more efficient and robust autonomous soaring in different environments.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Nos. 12202384 and U2241274), the Leading Talent Project for Scientific and Technological Innovation in Zhejiang Province (No. 2023R5220), and the Specialized Research Projects of Huanjiang Laboratory, China.

Author contributions

Siyuan ZHENG: investigation, methodology, data curation, visualization, and writing–original draft. Jiachi ZHAO: conceptualization, supervision, validation, and writing–review & editing. Lifang ZENG: supervision, project administration, and writing–review & editing. Zhouhong WANG: validation, methodology, and data curation. Jun LI: funding acquisition, supervision, and writing–review & editing.

Conflict of interest

Siyuan ZHENG, Jiachi ZHAO, Lifang ZENG, Zhouhong WANG, and Jun LI declare that they have no conflict of interest.

References

Allen M, 2005. Autonomous soaring for improved endurance of a small uninhabitated air vehicle. The 43rd AIAA Aerospace Sciences Meeting and Exhibit, article 1025. https://doi.org/10.2514/6.2005-1025

Chan WL, Lee CS, Hsiao FB, 2011. Real-time approaches to the estimation of local wind velocity for a fixed-wing unmanned air vehicle. *Measurement Science and Technology*, 22(10):105203. https://doi.org/10.1088/0957-0233/22/10/105203

Chung JJ, Lawrance NRJ, Sukkarieh S, 2015. Learning to soar: resource-constrained exploration in reinforcement learning. *International Journal of Robotics Research*, 34(2): 158-172. https://doi.org/10.1177/0278364914553683

Depenbusch NT, Bird JJ, Langelaan JW, 2018a. The AutoSOAR autonomous soaring aircraft part 2: hardware implementation and flight results. *Journal of Field Robotics*, 35(4): 435-458. https://doi.org/10.1002/rob.21747

Depenbusch NT, Bird JJ, Langelaan JW, 2018b. The AutoSOAR autonomous soaring aircraft, part 1: autonomy algorithms. *Journal of Field Robotics*, 35(6):868-889. https://doi.org/10.1002/rob.21782

Edwards DJ, Silverberg LM, 2010. Autonomous soaring: the montague cross-country challenge. *Journal of Aircraft*, 47(5):1763-1769. https://doi.org/10.2514/1.C000287

Edwards DJ, Kahn AD, Kelly M, et al., 2016. Maximizing net power in circular turns for solar and autonomous soaring aircraft. *Journal of Aircraft*, 53(5):1237-1247. https://doi.org/10.2514/1.C033634

Huey RB, Deutsch C, 2016. How frigate birds soar around the doldrums. *Science*, 353(6294):26-27. https://doi.org/10.1126/science.aag1865

Kahn AD, 2017. Atmospheric thermal location estimation. *Journal of Guidance, Control, and Dynamics*, 40(9):2363-2369. https://doi.org/10.2514/1.G002782

Langelaan JW, Alley N, Neidhoefer J, 2011. Wind field estimation for small unmanned aerial vehicles. *Journal of Guidance, Control, and Dynamics*, 34(4):1016-1030. https://doi.org/10.2514/1.52532

Lawrance NRJ, Sukkarieh S, 2011. Autonomous exploration of a wind field with a gliding aircraft. *Journal of Guidance, Control, and Dynamics*, 34(3):719-733. https://doi.org/10.2514/1.52236

Liu D, Lu F, Yang T, et al., 2021. A review of dynamic soaring: a new aapproach to extending the endurance of fixed-wing UAVs. Proceedings of the Unmanned Systems Summit, p.71-77. https://doi.org/10.26914/c.cnkihy.2021.020488

MacCready PB, 1958. Optimum airspeed selector. *Soaring*, 10(11):10.

Moore RJD, Thurrowgood S, Srinivasan MV, 2012. Vision-only estimation of wind field strength and direction from an aerial platform. IEEE/RSJ International Conference on Intelligent Robots and Systems, p.4544-4549. https://doi.org/10.1109/IROS.2012.6385682

Notter S, Zürn M, Groß P, et al., 2019. Reinforced learning to cross-country soar in the vertical plane of motion. AIAA Scitech Forum, article 1420. https://doi.org/10.2514/6.2019-1420

Notter S, Schimpf F, Fichter W, 2021. Hierarchical reinforcement learning approach towards autonomous cross-country soaring. AIAA Scitech Forum, article 2010.

https://doi.org/10.2514/6.2021-2010

Notter S, Müller G, Fichter W, 2022. Integrated updraft localization and exploitation: end-to-end type reinforcement learning approach. Proceedings of the 2022 CEAS EuroGNC conference, CEAS-GNC-2022-077.

Powers TC, Silverberg LM, Gopalarathnam A, 2020. Artificial lumbered flight for autonomous soaring. *Journal of Guidance, Control, and Dynamics*, 43(3):553-566. https://doi.org/10.2514/1.G004397

Reddy G, Celani A, Sejnowski TJ, et al., 2016. Learning to soar in turbulent environments. *Proceedings of the National Academy of Sciences of the United States of America*, 113(33):E4877-E4884. https://doi.org/10.1073/pnas.1606075113

Reddy G, Wong-Ng J, Celani A, et al., 2018. Glider soaring via reinforcement learning in the field. *Nature*, 562(7726): 236-239. https://doi.org/10.1038/s41586-018-0533-0

Rhudy MB, Larrabee T, Chao HY, et al., 2013. UAV attitude, heading, and wind estimation using GPS/INS and an air data system. AIAA Guidance, Navigation, and Control Conference, article 5201. https://doi.org/10.2514/6.2013-5201

Walton C, Kaminer I, Dobrokhodov V, et al., 2018. Alternate strategies for optimal unmanned aerial vehicle thermaling. *Journal of Aircraft*, 55(6):2347-2356. https://doi.org/10.2514/1.C035018

Wharington J, Herszberg I, 1998. Control of a high endurance unmanned air vehicle. Proceedings of the 21st ICAS Congress.

Woodbury TD, Dunn C, Valasek J, 2014. Autonomous soaring using reinforcement learning for trajectory generation. The 52nd Aerospace Sciences Meeting, article 0990. https://doi.org/10.2514/6.2014-0990

Zhao JC, Li J, Zeng LF, 2023. Energy-harvesting strategy investigation for glider autonomous soaring using reinforcement learning. *Aerospace*, 10(10):895. https://doi.org/10.3390/aerospace10100895