

# Reinforcement learning-based close formation control for underactuated surface vehicle with prescribed performance and time-varying state constraints

Huizi Chen<sup>a</sup>, Huaicheng Yan<sup>b</sup>, Yueying Wang<sup>a</sup>, Shaorong Xie<sup>a</sup>, Dan Zhang<sup>a,\*</sup>

<sup>a</sup> School of Mechanical Engineering and Automation, Shanghai University, Shanghai, 200444, China

<sup>b</sup> School of Information Science and Engineering, East China University of Science and Technology, Shanghai, 200237, China

## ARTICLE INFO

### Keywords:

Close formation control  
Finite-time sliding mode control  
Reinforcement learning  
Actor-critic neural network  
Prescribed formation performance  
Time-varying state constraints

## ABSTRACT

This paper studies close formation control problem with prescribed performance and time-varying state constraints for a group of 4-degrees-of-freedom (DOF) underactuated surface vehicles (USVs) subject to actuator faults, input saturation and input delay. A finite-time sliding mode control (SMC) scheme based on reinforcement learning (RL) algorithm is introduced to guarantee prescribed formation performance without violating velocity error constraints. By using actor-critic neural network (NN)-based RL algorithm, the actuator faults and system uncertainties are accurately estimated. Afterwards, an exponential decreasing boundary function is developed to suppress overshoot more reasonably, and a novel mechanism of switching gain is given to alleviate chattering inherent in SMC while the RL-based compensation term is constructed to handle the formation accuracy problem caused by the reduced switching gain. Besides, auxiliary nonlinear continuous function and Pade approximation have been successfully applied to process actuator saturation and input delay, respectively. Numerical simulations and experimental results are exhibited to verify the effectiveness and superior formation performance of the proposed control method.

## 1. Introduction

Formation control investigates the problem of multi-agent spatial distribution. Compared with a single USV, a group of USVs can perform a more complex task by maintaining a desired formation geometry (Chan et al., 2021). Currently, various control schemes are applied in formation control, including leader-follower strategy (Li et al., 2020b), event-triggered approach (Guo and Chen, 2020), virtual structures (Zhao et al., 2020) in either a centralized or decentralized manner. The most applicable one is the decentralized leader-follower strategy, which is simplicity, scalability and consumes less computation effort and communication resource (Zhao et al., 2015).

Distinguished from the above-mentioned formation, close formation generally refers to the formation with small relative distance among vehicles, which can reduce the wave-making resistance interference of adjacent vehicles and improve the shielding effect of surrounding vehicles under incoming water, so as to reduce the navigation resistance and increase the voyage distance (Liu, 2019). Hence, close formation is often applied in some special scenes, such as energy supply between

ships with non-zero speed, long distance formation navigation, etc. Since 2014, the Rim of the Pacific has taken close formation as a necessary training subject to improve the joint combat capability of various countries (Carter, 2018). In 2020, the "sea train" project is released by the Defense Advanced Research Projects Agency, and the close formation is used to reduce the collective wave-making resistance of USV as much as possible (Ellingsen, 2022).

During close formation, USVs will inevitably suffer from the coupled disturbances from surrounding vehicles and marine environment. Vigorous roll motion induced by the coupled disturbances is one of the dangerous phenomenon leading to operation interruption, instruments failure and even overturning of vehicle (Jin et al., 2020). Compared with general formation control, close formation control needs to consider not only surge, sway, yaw motion, but also the roll motion. Besides, the position drift caused by the coupled motion of rolling and other DOF motion results in formation error and collision. Therefore, roll stabilization control is always a challenging research topic in close formation (Liu et al., 2015; Chen et al., 2020).

Over the past years, many marine vehicles are equipped with anti-

\* Corresponding author.

E-mail address: [dan.zhang@shu.edu.cn](mailto:dan.zhang@shu.edu.cn) (D. Zhang).

rolling devices to suppress the roll motion, such as gyrostabilizers, anti-rolling tanks, moving weights and stabilizing fins (Awad et al., 2018; Hinostroza et al., 2015). Nevertheless, these methods require additional instruments and installation costs, and increase the complexity of the system. To address the aforementioned problems, rudder roll stabilization (RRS) strategy is employed, and rudder action is the only actuator to change the yaw motion and roll motion (Zhao et al., 2019). According to the roll motion and course deviation, rudder can be properly controlled to reduce the roll angle, and course will not change dramatically. In Zhao et al. (2020), an arctangent function-based control scheme was applied in RRS of ships. To against the bad tracking performance, RRS control technology was developed to cope with the large roll motion (Liu et al., 2020). Based on disturbance observer, RRS scheme was designed to reduce the roll angle and course error of ships (Liang and Wen, 2019). However, the above discussions do not take close formation control issue into account.

Besides, how to handle state constraints and obtain good transient performance is a very important consideration for most physical systems. The prescribed performance of formation error, namely, the maximum overshoot, amplitude and convergence rate of formation error are all less than the predefined boundaries, means excellent tracking performance (Yue et al., 2020; Dai et al., 2020). In Liang et al. (2020), a finite-time formation controller was adopted to ensure that multi-agent system synchronizes with prescribed performances. Using prescribed performance control methodology, decentralized adaptive formation controller (AFC) was obtained to guarantee collision avoidance and connectivity maintenance (He et al., 2019). On the other hand, state constraints of USVs are also research hotspot. Generally, barrier Lyapunov function (BLF) is always employed to solve state constraints. So far, multiple types of BLF are introduced to escape the infinite state, such as the tan-type BLF (Zhao et al., 2020c), the log-type BLF (Dong et al., 2020) and the integral-type BLF (Wang et al., 2019). A tangent BLF-based controller was applied to deal with time-varying output constraints (Zhao et al., 2020a). To cope with full state constraints, BLF-based fuzzy controller was used for a class of nontriangular form system (Zhang et al., 2020). In Li et al. (2020a), BLF-based tracking controller was proposed and time-varying state constraints were not violated. In fact, all the mentioned constraint strategies can be extended to formation.

Another significant issue associated with formation is to deal with the lumped uncertainties including actuator fault, system uncertainty. There are many ways to estimate the lumped uncertainties, such as NN approximation (Liu et al., 2015; Yue et al., 2020), fuzzy logic strategy (Zhao et al., 2020c), and disturbance observer (Liang and Wen, 2019). Although these strategies can approach the uncertainties, the bounded approximation error cannot be eliminated. In comparison with other control strategies, SMC is robust to the lumped uncertainties (Mobayen, 2016; Mobayen and Tchier, 2017). SMC can compensate bounded approximation error by choosing a switching gain whose amplitude is greater than the upper boundary of approximation error. Moreover, the amplitude of switching gain should not be too large, otherwise it will lead to system state chattering. Thus, SMC should reduce the gain to alleviate steady-state chattering when sliding variable enters the given vicinity (Lee et al., 2017). In Chen et al. (2020), a chattering suppression law was presented to handle excessive switching gain. Aiming at serious chattering problem inherent in SMC, an adaptation mechanism was designed to tune the amplitude of switching gain to avoid chattering (Baek et al., 2019). Although chattering has been alleviated by reducing switching gain, tracking accuracy will be seriously degraded when switching gain is too small to deal with tracking error (Xie et al., 2020). Hence, a compensation strategy should be developed to offset the lost tracking error and improve tracking accuracy.

RL algorithm is a kind of intelligent technology (Liu et al., 2022; Zhao et al., 2020c), which can optimize strategy continuously to maximize the cumulative reward or minimize the cost. For actor-critic NN-based RL algorithm, critic-NN is applied to assess system

performance (Wang et al., 2021), which is regarded as the reinforcement learning signal and guides the operation of actor-NN to optimize control input and improve system performance. In Wang et al. (2020b), actor-critic NN-based tracking controller was applied in a continuous-time nonlinear system and NN reconstruction error was effectively reduced. To solve the tracking problem of an elastic joint robot, RL controller with actor-critic NN architecture was employed to ameliorate tracking performance (Ouyang et al., 2020). In Zheng et al. (2020), an actor-critic NN-based control scheme was constructed for an underactuated marine vessel to obtain the prominent system performance.

Motivated by the optimal policy of RL, this paper presents a RL-based SMC scheme to overcome the drawbacks of SMC in dealing with small switching gain. The advantages of RL are used to compensate for the lost tracking error caused by the reduced switching gain and improve tracking accuracy. Subsequently, the proposed scheme is developed into a decentralized finite-time frame to achieve the close formation of USVs. The exponential decreasing boundary functions are designed to ensure collision avoidance and roll suppression. Additionally, an auxiliary nonlinear continuous function and Pade approximation have been successfully introduced to process actuator saturation and input delay, respectively. The RL-based sliding variables are constructed to meet state constraints and overcome external disturbances. Compared with the previous works, the contributions of this paper can be summarized as follows:

- Unlike most of the results in Awad et al. (2018) and Zhao et al. (2019a), rudder is the only input for sway-yaw-roll motion. The sway-yaw error subsystem stabilization and roll reduction in close formation can be achieved simultaneously through the same control input (e.g. a single rudder).
- A novel prescribed boundary function is developed, which helps USVs formation avoid overlarge initial error by determining the maximum convergence time.
- Compared with SMC strategies proposed in Zhao et al. (2019b) and Chen et al. (2020), the RL-based SMC control strategy is utilized to achieve chattering suppression while maintaining a high-precision formation performance. The adaptive law of switching gain is introduced to improve the chattering of output-state by reducing switching gain, and compensate the lumped uncertainties and approximation error. Meanwhile, the actor-critic NN-based RL algorithm is given to produce a compensation term to handle the formation accuracy problem caused by the reduced switching gain.

The remainder of this paper is as follows. Section 2 presents problem formulations and some preliminaries. Section 3 describes the design procedure of the proposed control method and analysis of the stability of error systems. Section 4 illustrates the results of conducted numerical simulations. Section 5 summarizes some conclusions and future works.

## 2. Problem formulation and preliminaries

### 2.1. Preliminaries

To achieve formation control objective, some existing lemmas will be used.

**Lemma 1.** (Liu et al., 2015). *The following NN technology is applied to approximate any continuous unknown function. The analytical expression of NN can be shown as follows*

$$f_n(Z_n) = W_n^{*T} H_n(Z_n) + \sigma_n \quad (1)$$

where  $Z_n \in R^n$  denotes input variable,  $H_n(Z_n) \in R^n$  is the basis function,  $\sigma_n$  is NN approximation error and  $|\sigma_n| \leq \bar{\sigma}_n$ ,  $\bar{\sigma}_n$  is a desired precision and  $\bar{\sigma}_n > 0$ ,  $W_n^* \in R^n$  is ideal weight, which can be shown as:

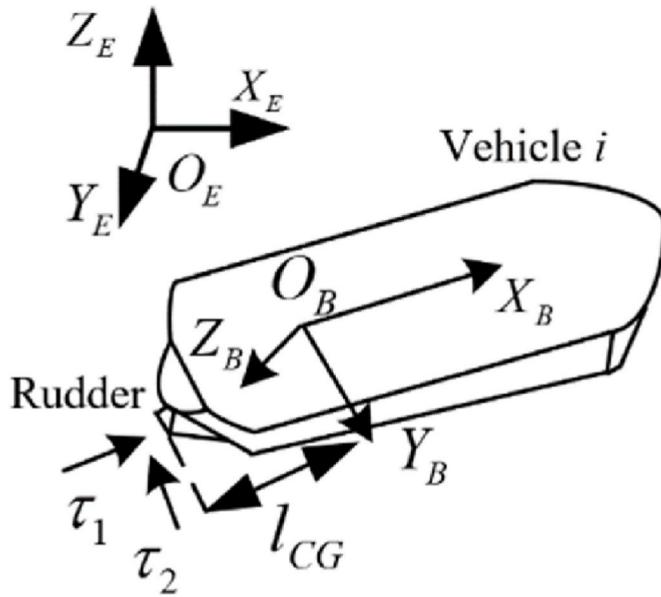


Fig. 1. 4-DOF plant of underactuated autonomous surface vehicle.

$W_n^* = \underset{\widehat{W}_n}{\operatorname{argmin}} \left\{ \sup_{Z_n \in \Omega} |f_n(Z_n) - \widehat{W}_n^T H_n(Z_n)| \right\}$  where  $\widehat{W}_n$  is the estimation of  $W_n^*$ .

**Lemma 2.** (Wang et al., 2017). For  $\omega_n \in R^n$  and  $0 < l < 1$ , the following equation holds

$$\left( \sum_{q=1}^n |\omega_q| \right)^l \leq \sum_{q=1}^n |\omega_q|^l \quad (2)$$

**Lemma 3.** (Wang et al., 2017). For variables  $\omega, \varsigma$  and positive constants  $\mu, \iota$ , the following equation can be obtained as

$$|\omega|^\mu |\varsigma|^\theta \leq \frac{\mu}{\mu + \theta} \iota |\omega|^{\mu+\theta} + \frac{\theta}{\mu + \theta} \iota^{-\frac{\mu}{\theta}} |\varsigma|^{\mu+\theta} \quad (3)$$

**Definition 1.** (Wang et al., 2017). The equilibrium  $\xi = 0$  of the nonlinear system  $\dot{\xi} = f(\xi, u)$  is semi-global practical finite-time stable (SGPFS) if for all  $\xi(t_0) = \xi_0$ , there exists  $\epsilon > 0$  and a settling time  $T(\epsilon, \xi_0) < \infty$  to make  $\|\xi(t)\| < \epsilon$ , for all  $t \geq t_0 + T$ .

**Lemma 4.** (Wang et al., 2017). Consider the nonlinear system  $\dot{\xi} = f(\xi, u)$ , for any scalars  $\alpha > 0, \beta > 0$  and  $0 < l < 1$ , the Lyapunov function of the system satisfies  $\dot{V}(x) \leq -\alpha V^l(x) + \beta$ , then this system is SGPFS.

**Lemma 5.** (Wang et al., 2018). The following systems

$$\begin{aligned} \dot{\hat{\xi}} &= \xi_u + f_u \\ \dot{\xi}_u &= -\nu_1 L^{1/2} \operatorname{sig}^{1/2}(\hat{\xi} - \xi) + \hat{D} \\ \hat{D} &= -\nu_2 L \operatorname{sgn}(\hat{D} - \xi_u) \end{aligned} \quad (4)$$

where  $L, \nu_1, \nu_2 > 0$  and  $\operatorname{sig}^\alpha(x) = |x|^\alpha \operatorname{sgn}(x)$ , and  $f_u$  is nonlinear term, the term  $\hat{D}$  can be exactly identified within a finite time.

## 2.2. Dynamic model of USV

A group of  $n$  USVs with 4-DOF motion are considered, including surge, sway, yaw and roll motion. The  $i$ th ( $i = 1, \dots, n$ ) 4-DOF plant of vehicle is shown in Fig. 1 and the corresponding kinematics can be

described as

$$\begin{aligned} \dot{x}_i &= u_i \cos \psi_i - v_i \cos \varphi_i \sin \psi_i \\ \dot{y}_i &= u_i \sin \psi_i + v_i \cos \varphi_i \cos \psi_i \\ \dot{\psi}_i &= r_i \cos \varphi_i \\ \dot{\varphi}_i &= p_i \end{aligned} \quad (5)$$

where  $x_i, y_i, \psi_i, \varphi_i$  are the positional variables in surge motion, sway motion, yaw angle and roll angle, respectively;  $u_i, v_i, r_i, p_i$  are surge velocity, sway velocity, yaw angular rate and roll angular rate, respectively.

For 4-DOF dynamics model of USV, propulsion force and torque are produced by the coupling of the rudder and thruster, and rudder is the only device used for yaw control and roll reduction. Kinetics of the  $i$ th vehicle can be written as (Jin et al., 2020)

$$\begin{aligned} \dot{u}_i &= \frac{m_{22i}}{m_{11i}} v_i r_i - \frac{d_{11i}}{m_{11i}} u_i^2 + \frac{\tau_{1i} + d_{ui}}{m_{11i}} \\ \dot{v}_i &= -\frac{m_{11i}}{m_{22i}} u_i r_i - \frac{d_{22i}}{m_{22i}} v_i + \frac{\tau_{2i} + d_{vi}}{m_{22i}} \\ \dot{r}_i &= \frac{m_{11i} - m_{22i}}{m_{33i}} u_i v_i - \frac{d_{33i}}{m_{33i}} r_i + \frac{\tau_{3i} l_{CG} + d_{ri}}{m_{33i}} \\ \dot{p}_i &= -\frac{d_{44i}}{m_{44i}} p_i - l_{pi} \varphi_i + \frac{\tau_{4i} l_{zi} + d_{pi}}{m_{44i}} \end{aligned} \quad (6)$$

where  $\tau_{1i} = \operatorname{sat}(\tau_{1i}^F(t - \varepsilon_t))$ ,  $\tau_{2i} = \operatorname{sat}(\tau_{2i}^F(t - \varepsilon_t))$  denote the surge force and lift force made by the surface of the rudder under the input saturation, time delay and actuator faults, and  $\operatorname{sat}(\cdot)$  is saturation function, and  $\varepsilon_t > 0$  is the unknown input delay;  $l_{zi}$  is the distance of rudder roll suppression and  $l_{CG}$  is the distance from center of gravity to rudder stock;  $d_{ui}, d_{vi}, d_{ri}, d_{pi}$  are the unknown and time-varying external disturbances from marine environment and adjacent vehicles;  $m_{11i}, m_{22i}, m_{33i}, m_{44i}$  are inertia coefficients;  $d_{11i}, d_{22i}, d_{33i}, d_{44i}$  are hydrodynamic damping coefficients;  $l_{pi}$  is the roll factor of vehicle.

**Remark 1.** The purpose of this paper is to study the close formation control problem of USVs subject to the roll motion, actuator faults, input saturation and input delay. Roll motion is common in vehicle cruising. However, large roll motion will degrade the tracking performance of vehicle, damage instruments, and sometimes may cause vehicle to overturn. Therefore, it is reasonable to consider roll motion. Additionally, vehicle is driven by the rudder and thruster. The actuator faults, input saturation and input delay of the rudder and thruster are inevitable in time-varying and complex marine engineering. From a practical point of view, dynamics model (6) is effective and reasonable.

**Assumption 1.** (Zhao et al., 2019b). The external disturbances  $d_{ui}, d_{vi}, d_{ri}, d_{pi}$  are time-varying and bounded.

**Assumption 2.** (Zhao et al., 2019b). The reference trajectories  $x_{i-1}, \dot{x}_{i-1}, \ddot{x}_{i-1}, y_{i-1}, \dot{y}_{i-1}, \ddot{y}_{i-1}, \psi_{i-1}, \dot{\psi}_{i-1}, \ddot{\psi}_{i-1}$  are smooth and bounded. The roll angle of USV is bounded and there exists  $|\varphi_i| \leq \varphi_{\max} < \frac{\pi}{2}$  with  $\varphi_{\max}$  being a positive constant.

**Remark 2.** The coupled disturbances force or moment are caused by ocean and adjacent hull, which are often considered as bounded, low-frequency and time-varying with finite energy, so it is generally regarded as time-varying and bounded signals. Thus, Assumption 1 is reasonable. For USV  $i$ -1, its trajectory is taken as the desired trajectory by USV  $i$ , and it cannot be too large. Once the state variables  $x_{i-1}, \dot{x}_{i-1}, \ddot{x}_{i-1}, y_{i-1}, \dot{y}_{i-1}, \ddot{y}_{i-1}, \psi_{i-1}, \dot{\psi}_{i-1}, \ddot{\psi}_{i-1}$  are arbitrarily large, the actuator cannot supply enough control input, resulting in formation error. From the perspective of controllability for USV, Assumption 2 is reasonable.

In marine engineering, the failures, saturation and delay of actuators are common in thruster and rudder for vehicles. Considering the above shortcomings, actual control input can be mathematically modeled as (Liu et al., 2019)

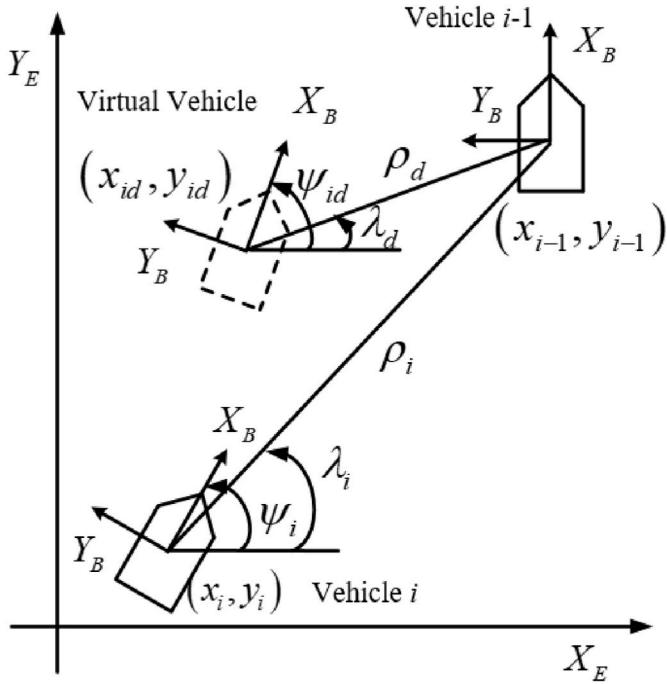


Fig. 2. Leader-follower formation.

$$\text{sat}(\tau_{ki}^F(t - \varepsilon_t)) = \varpi_{ki} \text{sat}(\tau_{ki}^c(t - \varepsilon_t)) + \bar{\tau}_{ki}(t - \varepsilon_t), k = 1, 2 \quad (7)$$

where  $\varpi_{ki}$  represents the efficiency coefficient of unknown time-varying actuator,  $\bar{\tau}_{ki}(t - \varepsilon_t)$  is additive faults,  $\tau_{ki}^c(t - \varepsilon_t)$  stands for the commanded control input.

Herein, the case  $\varpi_{ki} = 1$  and  $\bar{\tau}_{ki}(t - \varepsilon_t) = 0$  indicates that the  $k$ th actuator is not faulty, while  $\varpi_{ki} = 0$  represents that the  $k$ th actuator does not provide force. Besides,  $\varpi_{ki} \in (0, 1)$  means that the  $k$ th actuator loses its effectiveness partially.

**Assumption 3.** (Liu et al., 2019). There is a boundary  $\tau_{kM}$  such that  $\tau_{kM} \geq \text{sat}(\tau_{ki}^c(t - \varepsilon_t))$ .

**Assumption 4.** (Liu et al., 2019). The additive fault  $\bar{\tau}_{ki}(t - \varepsilon_t)$  is unknown but bounded as  $\|\bar{\tau}_{ki}(t - \varepsilon_t)\| \leq \bar{\tau}_{kM}$ , and  $\bar{\tau}_{kM}$  is a positive constant.

To tackle the problem of time-delay in  $\tau_{ki}^c(t - \varepsilon_t)$ , Pade approximation strategy and Laplace transform are employed to obtain control input without time delay.

$$L[\text{sat}(\tau_{ki}^c(t - \varepsilon_t))] = \exp(-\varepsilon_t s) L[\text{sat}(\tau_{ki}^c(t))] \approx \frac{1 - \varepsilon_t s/2}{1 + \varepsilon_t s/2} L[\text{sat}(\tau_{ki}^c(t))] \quad (8)$$

where  $s$  is the Laplace variable,  $L[\cdot]$  denotes the Laplace transform,  $\exp(\cdot)$  is exponential function.

To compensate time delay of actuator, an auxiliary variable  $\chi_{ki}(t)$  is defined as

$$L[\chi_{ki}(t)] = \frac{1 - \varepsilon_t s/2}{1 + \varepsilon_t s/2} L[\text{sat}(\tau_{ki}^c(t))] + L[\text{sat}(\tau_{ki}^c(t))] \quad (9)$$

Then, equation (9) can also be described as

$$2L[\text{sat}(\tau_{ki}^c(t))] = L[\chi_{ki}(t)] + \frac{\varepsilon_t s}{2} L[\chi_{ki}(t)] \quad (10)$$

Assuming that the initial condition is  $\chi_{ki}(0) = 0$ , using inverse the Laplace transform, equation (10) becomes

$$\dot{\chi}_{ki}(t) = \frac{4}{\varepsilon_t} \text{sat}(\tau_{ki}^c(t)) - \frac{2}{\varepsilon_t} \chi_{ki}(t) \quad (11)$$

Additionally, according to equation (8), equation (9) in time domain

can be expressed as

$$\text{sat}(\tau_{ki}^c(t - \varepsilon_t)) = \chi_{ki}(t) - \text{sat}(\tau_{ki}^c(t)) \quad (12)$$

Substituting equation (12) into (7), equation (7) can be rewritten as

$$\begin{aligned} \text{sat}(\tau_{ki}^F(t - \varepsilon_t)) &= \chi_{ki}(t) - \text{sat}(\tau_{ki}^c(t)) + (\varpi_{ki} - 1)(\chi_{ki}(t) - \text{sat}(\tau_{ki}^c(t))) \\ &\quad + \bar{\tau}_{ki}(t - \varepsilon_t) \end{aligned} \quad (13)$$

To remove the sharp point of saturation function at  $|\tau_{ki}^c(t)| = \tau_{kM}$ , smooth continuous function  $f(\tau_{ki}^c(t))$  is utilized to replace  $\text{sat}(\tau_{ki}^c(t))$  and shows as follow

$$f(\tau_{ki}^c(t)) = \tau_{kM} \tanh(\tau_{ki}^c(t)) \quad (14)$$

Function error between saturated control input (7) and function (14) is defined as

$$\Delta \tau_{ki}^c = \text{sat}(\tau_{ki}^c(t)) - f(\tau_{ki}^c(t)) \quad (15)$$

Based on assumption 3 and equation (14), the boundary of function error (15) can be expressed as

$$|\Delta \tau_{ki}^c| = |\text{sat}(\tau_{ki}^c(t)) - f(\tau_{ki}^c(t))| \leq \tau_{kM} (1 - \tanh(1)) \quad (16)$$

**Remark 3.** In contrast with the published references (Wang et al., 2020a), which handles input saturation by using the complicated anti-windup compensator, a simpler auxiliary nonlinear continuous function (14) is applied to address this problem.

By introducing equation (15) into (13), we obtain

$$\begin{aligned} \text{sat}(\tau_{ki}^F(t - \varepsilon_t)) &= \chi_{ki}(t) - (\Delta \tau_{ki}^c + f(\tau_{ki}^c(t))) \\ &\quad + (\varpi_{ki} - 1)[\chi_{ki}(t) - (\Delta \tau_{ki}^c + f(\tau_{ki}^c(t)))] + \bar{\tau}_{ki}(t - \varepsilon_t) \end{aligned} \quad (17)$$

To simplify analytic expression, the variables  $\tau_{ki}^c(t), f(\tau_{ki}^c(t)), \chi_{ki}(t), \bar{\tau}_{ki}(t - \varepsilon_t)$  will be described as  $\tau_{ki}^c, f(\tau_{ki}^c), \chi_{ki}, \bar{\tau}_{ki}$  in following text.

Define the following equation as

$$\begin{aligned} f_{ui} &= \frac{m_{22i}}{m_{11i}} v_i r_i - \frac{d_{11i}}{m_{11i}} u_i^2 + \frac{(\varpi_{1i} - 1)[\chi_{1i} - (\Delta \tau_{1i}^c + f(\tau_{1i}^c))] + \bar{\tau}_{1i}}{m_{11i}} \\ f_{vi} &= -\frac{m_{11i}}{m_{22i}} v_i r_i - \frac{d_{22i}}{m_{22i}} v_i + \frac{(\varpi_{2i} - 1)[\chi_{2i} - (\Delta \tau_{2i}^c + f(\tau_{2i}^c))] + \bar{\tau}_{2i}}{m_{22i}} \\ f_{ri} &= \frac{m_{11i} - m_{22i}}{m_{33i}} u_i v_i - \frac{d_{33i}}{m_{33i}} r_i + \frac{(\varpi_{2i} - 1)[\chi_{2i} - (\Delta \tau_{2i}^c + f(\tau_{2i}^c))] + \bar{\tau}_{2i}}{m_{33i}} l_{CG} \\ f_{pi} &= -\frac{d_{44i}}{m_{44i}} p_i - l_{pi} \varphi_i + \frac{(\varpi_{2i} - 1)[\chi_{2i} - (\Delta \tau_{2i}^c + f(\tau_{2i}^c))] + \bar{\tau}_{2i}}{m_{44i}} l_{zi} \end{aligned} \quad (18)$$

Under assumptions (1)–(4), the lumped uncertainties  $f_{ui}, f_{vi}, f_{ri}, f_{pi}$  including actuator fault and system uncertainty are time-varying and bounded. Based on equations (11), (17) and (18), the plant (6) can be rewritten as follows

$$\begin{aligned} \dot{u}_i &= f_{ui} + \frac{\chi_{1i} - (\Delta \tau_{1i}^c + f(\tau_{1i}^c))}{m_{11i}} \\ \dot{v}_i &= f_{vi} + \frac{\chi_{2i} - (\Delta \tau_{2i}^c + f(\tau_{2i}^c))}{m_{22i}} \\ \dot{r}_i &= f_{ri} + \frac{[\chi_{2i} - (\Delta \tau_{2i}^c + f(\tau_{2i}^c))] l_{CG}}{m_{33i}} \\ \dot{p}_i &= f_{pi} + \frac{[\chi_{2i} - (\Delta \tau_{2i}^c + f(\tau_{2i}^c))] l_{zi}}{m_{44i}} \\ \dot{\chi}_{ki} &= \frac{4}{\varepsilon_t} (\Delta \tau_{ki}^c + f(\tau_{ki}^c)) - \frac{2}{\varepsilon_t} \chi_{ki} \end{aligned} \quad (19)$$

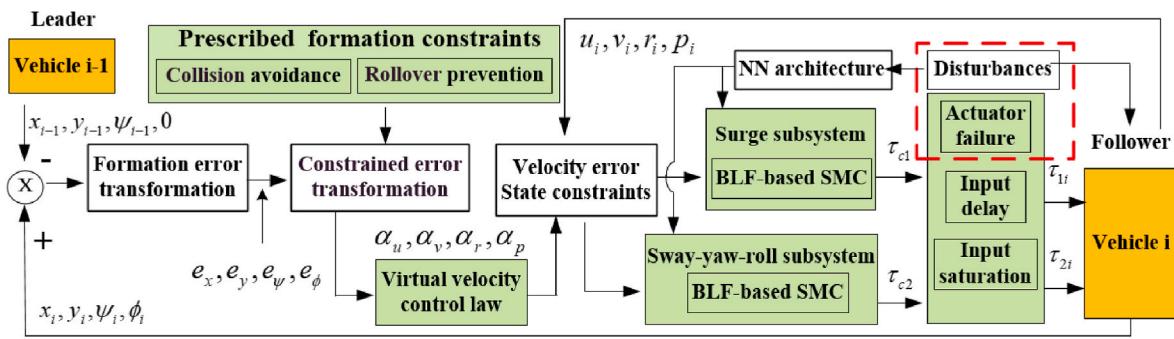


Fig. 3. Control architecture for USVs formation.

### 2.3. Decentralized leader-follower formation

A decentralized formation strategy is utilized to force a group of USVs to achieve leader-follower formation. A virtual vehicle is assumed to maintain the desired formation with leader vehicle  $i-1$ . And if the follower  $i$  can track the virtual vehicle, the desired formation can be formed. Every vehicle  $i$  only receives state information from its immediate leader vehicle  $i-1$ . The state information of global leader is derived from predetermined reference trajectory. Therefore, the communication relationship of each vehicle forms a static and simple directed spanning tree. The position about two vehicles in leader-follower formation is presented in Fig. 2. The line-of-sight (LOS) range  $\rho_i$  and angle  $\lambda_i$  between vehicle  $i$  and vehicle  $i-1$  are defined as

$$\begin{aligned} \rho_i &= \sqrt{(x_{i-1} - x_i)^2 + (y_{i-1} - y_i)^2} \\ \lambda_i &= \arctan(y_{i-1} - y_i, x_{i-1} - x_i) \end{aligned} \quad (20)$$

According to the formation mechanism, the path of virtual vehicle is defined as

$$\begin{pmatrix} x_{id} \\ y_{id} \\ \psi_{id} \end{pmatrix} = \begin{pmatrix} x_{i-1} \\ y_{i-1} \\ \psi_{i-1} \end{pmatrix} + \begin{pmatrix} \cos \psi_{i-1} & -\sin \psi_{i-1} & 0 \\ \sin \psi_{i-1} & \cos \psi_{i-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \rho_d \cos \lambda_d \\ \rho_d \sin \lambda_d \\ 0 \end{pmatrix} \quad (21)$$

where  $(x_{id}, y_{id})$  is the position of virtual vehicle, which represents the position that follower needs to track to maintain the desired formation,  $(\rho_d, \lambda_d)$  is the desired LOS range and angle.

According to the leader vehicle state information  $x_{i-1}, y_{i-1}, \psi_{i-1}$  and the desired formation information  $\rho_d, \lambda_d$ , using equation (21),  $x_i = x_{id}$ ,  $y_i = y_{id}$ ,  $\psi_i = \psi_{id}$  can be obtained, which is equivalent to  $\rho_i = \rho_d$ ,  $\lambda_i = \lambda_d$ , and it means that the desired formation control is achieved. Therefore, the formation problem of leader and followers can be transformed into the path following of followers.

Define the following relative distance and angle errors

$$\begin{aligned} e_x &= (x_i - x_{id})\cos \psi_i + (y_i - y_{id})\sin \psi_i \\ e_y &= -(x_i - x_{id})\sin \psi_i + (y_i - y_{id})\cos \psi_i \\ e_\psi &= \psi_i - \psi_{id} \\ e_\phi &= \phi_i - \phi_d \end{aligned} \quad (22)$$

where  $\phi_d$  represents a desired roll angle with  $\phi_d = 0$ .

For close formation, the collision between adjacent USVs is easy to occur due to small formation distance, and the roll motion caused by adjacent USVs and surrounding environment cannot be ignored. Thus, the following constraints are designed to avoid the collision and severe roll motion of vehicles.

$$-\underline{e}_j < e_j < \bar{e}_j \quad (23)$$

where  $j = x, y, \psi, \phi$ ,  $\underline{e}_j$  and  $\bar{e}_j$  are the predefined lower and upper boundary of formation error, respectively.

Considering convergence rate, the maximum overshoot, the maximum steady-state, and terminal time into formation error, the

exponential decreasing boundary functions  $\underline{e}_j, \bar{e}_j$  are designed to ensure prescribed performance.

$$\begin{aligned} \bar{e}_j(t) &= \begin{cases} (\bar{e}_{j,0} - \bar{e}_{j,t_f})\exp\left(-\bar{k}_j \tan\left(\frac{\pi t}{2t_f}\right)\right) + \bar{e}_{j,t_f}, & 0 \leq t < t_f \\ \bar{e}_{j,t_f}, & t \geq t_f \end{cases} \\ \underline{e}_j(t) &= \begin{cases} (\underline{e}_{j,0} - \underline{e}_{j,t_f})\exp\left(-\bar{k}_j \tan\left(\frac{\pi t}{2t_f}\right)\right) + \underline{e}_{j,t_f}, & 0 \leq t < t_f \\ \underline{e}_{j,t_f}, & t \geq t_f \end{cases} \end{aligned} \quad (24)$$

where  $\underline{e}_{j,0}$  and  $\bar{e}_{j,0}$  are initial states,  $\underline{e}_{j,t_f}$  and  $\bar{e}_{j,t_f}$  are the steady-state error after system error converges in a predefined time  $t_f$ ,  $\underline{k}_j$  and  $\bar{k}_j$  are the convergence rates.

Different from traditional boundary function, function (24) has the following important features.

- 1) The terminal time  $t_f$  of error convergence can be predefined.
- 2) Once the terminal time is determined, the convergent rate of performance function (24) can be regulated by choosing appropriate parameters  $\bar{k}_j, \underline{k}_j$ , and initial overshoot is also avoided.

**Remark 4.** The convergence rate of the exponential performance function depends on the parameter  $\bar{k}_j, \underline{k}_j$ . Therefore, the exponential performance function avoids excessive control input by reducing the initial convergence rate of system, but it will destroy the transient performance of system. However, new performance function (24) can obtain different convergence rates by setting terminal time  $t_f$  and selecting appropriate parameters  $\bar{k}_j, \underline{k}_j$ , which can significantly improve the initial control input and reduce the formation error caused by overshoot.

For kinematics (5) and kinetics (19), a decentralized controller will be devised to achieve the following formation control objective:

- 1) During close formation, each vehicle follows its immediate leader without violating the constraints given in (23).
- 2) All the variables are uniformly bounded, and formation errors  $e_x, e_y, e_\psi, e_\phi$  converge to the neighborhood near zero in finite-time.

### 3. Controller design

In this part, the formation control architecture is elaborately established, as shown in Fig. 3. According to the kinematic information of leader-follower and the formation error transformation mechanism in (20), formation error (22) is obtained. Afterwards, by using backstepping technology and constrained error transformation strategy, the

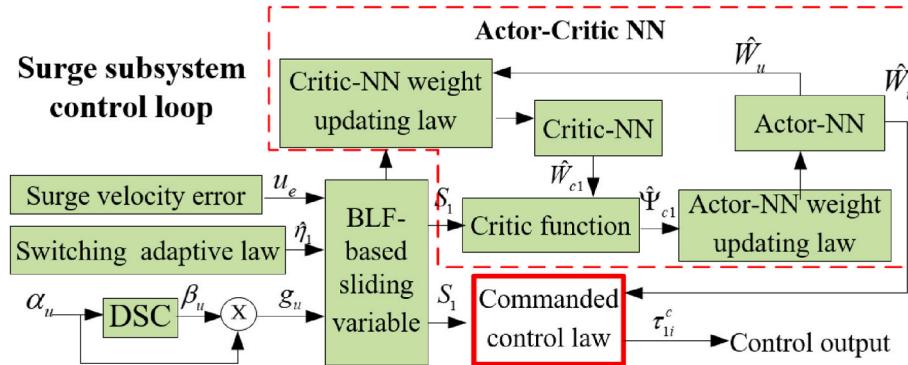


Fig. 4. Signal flow of the proposed surge controller.

virtual velocity control law is proposed to maintain formation error within prescribed boundary constraints. Furthermore, the RL-based SMC is designed in terms of velocity error constraints and actor-critic NN architecture. Eventually, formation control objective is achieved for a team of USVs with actuator failure, input delay, input saturation and disturbances.

### 3.1. Constrained error transformation

To meet performance constraints, an error transformation function is defined as

$$z_j = \frac{e_j}{\bar{e}_j - e_j} q(e_j) + \frac{e_j}{\bar{e}_j + e_j} (1 - q(e_j)) \quad (25)$$

where parameter  $q(e_j)$  satisfies the following equation  $q(e_j) = \begin{cases} 1, & e_j \geq 0 \\ 0, & e_j < 0 \end{cases}$ .

**Remark 5.** In Dong et al. (2020), error transformation function was logarithmic function, which would cause the potential singularity problem of control law. Thus, nonlogarithmic error mapping function (25) is applied to address the singularity problem. Meanwhile, variable  $z_j$  is a continuous derivable function compared with Zhang et al. (2020c).

Based on equation (25), the derivatives of  $z_j$  are

$$\dot{z}_j = G_j \dot{e}_j - Q_j \quad (26)$$

$$\text{where } G_j = \frac{q(e_j)\bar{e}_j}{(\bar{e}_j - e_j)^2} + \frac{(1-q(e_j))\bar{e}_j}{(\bar{e}_j + e_j)^2}, Q_j = \frac{q(e_j)\dot{e}_j e_j}{(\bar{e}_j - e_j)^2} + \frac{(1-q(e_j))\dot{e}_j e_j}{(\bar{e}_j + e_j)^2}.$$

### 3.2. Kinematics control design

A close formation strategy is proposed to achieve the formation control objective, and the virtual velocity control law will be given to stabilize the kinematic tracking error.

Based on equations (5) and (22), the derivatives of  $e_x, e_y, e_\psi, e_\varphi$  are

$$\begin{aligned} \dot{e}_x &= u_i - \sqrt{\dot{x}_{id}^2 + \dot{y}_{id}^2} \cos e_\psi + r_i e_y \\ \dot{e}_y &= v_i + \sqrt{\dot{x}_{id}^2 + \dot{y}_{id}^2} \sin e_\psi - r_i e_x \\ \dot{e}_\psi &= r_i - \dot{\psi}_{id} \\ \dot{e}_\varphi &= p_i \end{aligned} \quad (27)$$

To promote the convergence of formation errors, the virtual velocity control laws are designed as

$$\begin{aligned} \alpha_u &= \sqrt{\dot{x}_{id}^2 + \dot{y}_{id}^2} \cos e_\psi - r_i e_y + G_x^{-1}(-k_{zx} z_x + Q_x) \\ \alpha_v &= -\sqrt{\dot{x}_{id}^2 + \dot{y}_{id}^2} \sin e_\psi + r_i e_x + G_y^{-1}(-k_{zy} z_y + Q_y) \\ \alpha_r &= \dot{\psi}_{id} + G_\psi^{-1}(-k_{z\psi} z_\psi + Q_\psi) \\ \alpha_p &= G_\varphi^{-1}(-k_{z\varphi} z_\varphi + Q_\varphi) \end{aligned} \quad (28)$$

where  $k_{zx}, k_{zy}, k_{z\psi}, k_{z\varphi}$  are positive constants.

When  $\alpha_u \rightarrow u_i, \alpha_v \rightarrow v_i, \alpha_r \rightarrow r_i, \alpha_p \rightarrow p_i$ , equation (28) can be further written as

$$\begin{aligned} \dot{e}_x &= G_x^{-1}(-k_{zx} z_x + Q_x) \\ \dot{e}_y &= G_y^{-1}(-k_{zy} z_y + Q_y) \\ \dot{e}_\psi &= G_\psi^{-1}(-k_{z\psi} z_\psi + Q_\psi) \\ \dot{e}_\varphi &= G_\varphi^{-1}(-k_{z\varphi} z_\varphi + Q_\varphi) \end{aligned} \quad (29)$$

Combining equation (25) with (29), the following equation can be obtained

$$\dot{z}_x = -k_{zx} z_x, \dot{z}_y = -k_{zy} z_y, \dot{z}_\psi = -k_{z\psi} z_\psi, \dot{z}_\varphi = -k_{z\varphi} z_\varphi \quad (30)$$

Choose the following Lyapunov function candidate

$$V_1 = \frac{1}{2} z_x^2 + \frac{1}{2} z_y^2 + \frac{1}{2} z_\psi^2 + \frac{1}{2} z_\varphi^2 \quad (31)$$

Taking time derivative of  $V_1$  along with (30), we have

$$\dot{V}_1 = -k_{zx} z_x^2 - k_{zy} z_y^2 - k_{z\psi} z_\psi^2 - k_{z\varphi} z_\varphi^2 \leq 0 \quad (32)$$

Therefore, once  $u_i - \alpha_u = 0, v_i - \alpha_v = 0, r_i - \alpha_r = 0, p_i - \alpha_p = 0$ , the prescribed formation errors  $z_x, z_y, z_\psi, z_\varphi$  will converge to zero. In next section, the target is to stabilize the dynamic error.

### 3.3. Kinetics control design with actor-critic NN

Surge controller and sway-yaw-roll controller are introduced in step 1 and step 2, which can promote the convergence of the kinetics tracking error in finite-time.

Define the velocity tracking errors as

$$\begin{aligned} u_e &= u_i - \alpha_u + \frac{\chi_{1i}}{2} - \xi_1 \\ v_e &= v_i - \alpha_v \\ r_e &= r_i - \alpha_r + \frac{\chi_{2i} l_{CG}}{2} - \xi_2 \\ p_e &= p_i - \alpha_p \end{aligned} \quad (33)$$

where  $\xi_1$  is an intermediate variable used to compensate for the function error between saturated control input  $\text{sat}(\tau_{ki}^c)$  and function (14),  $\xi_k = -\xi_k + \frac{1}{m_{kki}}(f(\tau_{ki}^c) - \tau_{ki}^c), k = 1, 2$ .

**Step 1.** In order to better achieve the surge controller design, the

signal flow of the controller is given in Fig. 4.

Differentiating  $u_e$ , we have

$$\dot{u}_e = \dot{u}_i - \dot{\alpha}_u + \frac{\dot{\chi}_{1i}}{2} - \dot{\xi}_1 \quad (34)$$

To avoid the computational expansion of  $\dot{\alpha}_u$ , DSC technology is introduced as follow

$$\varepsilon_u \dot{\beta}_u + \beta_u = \alpha_u, \beta_u(0) = \alpha_u(0) \quad (35)$$

where  $\varepsilon_u$  is a positive constant,  $\beta_u$  is the filter variable,  $\beta_u(0), \alpha_u(0)$  are initial conditions of  $\beta_u, \alpha_u$ .

Based on equations (19), (34) and (35), we have

$$\dot{u}_e = f_{ui} - \dot{\beta}_u + \xi_1 + \frac{1}{m_{11i}} (\tau_{1i}^c + \Delta \tau_{1i}^c) \quad (36)$$

Define filter error as

$$g_u = \beta_u - \alpha_u \quad (37)$$

According to equations (35) and (37), the variable  $\dot{\beta}_u$  can be described as

$$\dot{\beta}_u = -\frac{g_u}{\varepsilon_u} + \dot{\alpha}_u = -\frac{g_u}{\varepsilon_u} + B_u(\bullet) \quad (38)$$

where  $B_u(\bullet)$  is the time derivative of  $\alpha_u$ . Based on assumption 2, there exists a positive constant  $B_{uM}$  and  $|B_u| \leq B_{uM}$ .

To satisfy the surge velocity constraint and improve the transient performance and robustness in surge subsystem, the sliding variable based on logarithmic function is defined as

$$S_1 = \frac{1}{2} \ln \frac{k_{u_e}^2}{k_{u_e}^2 - (u_e + k_u)^2} + c_1 g_u + (\varepsilon_{f_u} - \hat{\eta}_1) \quad (39)$$

where  $k_u$  is a design parameter and it helps the control law avoid singularity,  $c_1$  is a design parameter,  $\varepsilon_{f_u}$  denotes the upper boundary of approximation error,  $c_1, \varepsilon_{f_u}$  are positive constants,  $\hat{\eta}_1$  is a positive compensation law of approximation error,  $k_{u_e}$  is continuous time-varying functions, which denote the upper boundary of variable  $u_e$ .

**Remark 6.** To enhance the robustness of vehicles against to lumped uncertainties, input saturation and time-delay, the novel sliding variable (39) is devised in terms of the velocity errors (33), filter errors (37), input error (15) and approximate errors. In contrast with the previous works in Zhang et al. (2020c) and Wang et al. (2019), which handles the problem of constant state constraints by establishing BLF, time-varying state constraints are applied in sliding variable (39) to meet different practical requirements.

According to equations (19), (34) and (37), the time derivative of (39) is

$$\dot{S}_1 = T_u \left[ f_{ui} + \xi_1 + \frac{1}{m_{11i}} (\tau_{1i}^c + \Delta \tau_{1i}^c + d_{ui}) \right] - T_{k_u} - (T_u + c_1) \dot{\beta}_u + c_1 \dot{\alpha}_u - \dot{\hat{\eta}}_1 \quad (40)$$

$$\text{where } T_u = \frac{u_e + k_u}{k_{u_e}^2 - (u_e + k_u)^2}, T_{k_u} = \frac{k_{u_e} (u_e + k_u)^2}{k_{u_e} [k_{u_e}^2 - (u_e + k_u)^2]}.$$

Based on lemma 1, actor-NN architecture can be used to approximate  $f_{ui}$ .

$$f_{ui} = W_u^{*T} H_u(Z) + \sigma_u \quad (41)$$

where  $W_u^{*T}$  is the optimal actor-NN weight,  $\sigma_u$  is NN approximation error and  $|\sigma_u| \leq \bar{\sigma}_u$ ,  $\bar{\sigma}_u$  denotes a desired precision and  $\bar{\sigma}_u > 0$ .

By utilizing equation (41), the following approximation equation can be obtained

$$f_{ui} = \widehat{W}_u^T H_u(Z) + \widetilde{W}_u^T H_u(Z) + \sigma_u \quad (42)$$

where  $\widehat{W}_u^T$  is the estimation of  $W_u^{*T}$ ,  $\widetilde{W}_u^T = W_u^{*T} - \widehat{W}_u^T$ .

Inspired by the adaptive dynamic programming method (Li et al., 2022), this paper utilizes the RL algorithm based on actor-critic NN to approximate the loss function in formation control, so as to obtain the optimal loss function and improve control output. To achieve the optimization of reward/penalty, the critic NN should be used to approximate the optimal loss function, evaluate the system performance and adjust the control output of actor NN. Therefore, the loss function based on critic NN, namely critic function, should be proposed.

To optimize formation performance continuously, the critic function is defined as

$$\Psi_{cl} = L_1 R_1 + |L_1 R_1| W_{cl}^T H_{cl}(Z) \quad (43)$$

The critic function includes primary critic signal and secondary critic signal. The primary critic signal reflects the current formation performance. Since the update law of critic NN is directly affected by actor NN, the secondary critic signal shows the formation performance after the control output is optimized.

Here, the critic signal vector  $R_1$  is chosen as

$$R_1 = \frac{L_2}{1 + \exp(-L_3 S_1)} - \frac{L_2}{1 + \exp(L_3 S_1)} \quad (44)$$

where  $L_2, L_3$  are positive constants,  $R_1 \in [-L_2, L_2]$  increases with the increase of sliding variable  $S_1$  and it corresponds to the formation errors performance,  $W_{cl}^T$  is the ideal weight of the critic-NN,  $H_{cl}(Z)$  is the basis function.

For equation (43), the actual critic function is

$$\widehat{\Psi}_{cl} = L_1 R_1 + |L_1 R_1| \widehat{W}_{cl}^T H_{cl}(Z) \quad (45)$$

where  $\widehat{W}_{cl}^T$  is the estimation of  $W_{cl}^T$ .

**Remark 7.** From critic function (45) and critic signal vector  $R_1$ ,  $\Psi_{cl}^T \Psi_{cl}$  increases with the increase of sliding variable  $S_1$ , vice versa. It indicates that  $\Psi_{cl}^T \Psi_{cl}$  can be regarded as the critic index of formation performance. Thus, actual critic function  $\widehat{\Psi}_{cl}$  will be applied to update actor-NN and construct the control policy to optimize formation performance.

Based on equations (40)-(45), the control law  $\tau_{1i}^c$  can be chosen as

$$\begin{aligned} \tau_{1i}^c &= \frac{m_{11i}}{T_u} [T_{k_u} + (T_u + c_1) \dot{\beta}_u - c_1 B_{uM} + \dot{\hat{\eta}}_1 + \tau_{1i}^{com} + \mu_1 + \tau_{1swi}] \\ &\quad - m_{11i} (\widehat{W}_u^T H_u(Z) + \xi_1) - \Delta \tau_{1i}^c - \widehat{d}_{ui} \tau_{1i}^{com} = -\frac{k_{c1}}{\exp(\hat{\eta}_1)} \widehat{\Psi}_{cl} \mu_1 \\ &= -|L_1 R_1| \left( 2\varphi_1 + \frac{k_{c1}}{\exp(\hat{\eta}_1)} \right) (1 + \varphi_1) - T_u \bar{\sigma}_u \end{aligned} \quad (46)$$

where  $\tau_{1i}^{com}$  is a compensation term to handle the problem that the reduced switching gain is insufficient to offset the lumped uncertainties and approximation errors,  $\mu_1$  is an additional control term to help reduce the reconstruction errors of actor-critic-NNs,  $\tau_{1swi}$  is the switching law and  $\tau_{1swi} = -\hat{\eta}_1 \text{sgn}(S_1) - k_{1sw} S_1$ ,  $k_{c1}$  is a positive constant,  $\varphi_1$  is a positive design parameter,  $\widehat{d}_{ui}$  is the estimation of  $d_{ui}$  and obtained from lemma 5.

To attenuate chattering in equation (46), the following adaptive law of switching gain is chosen as

$$\dot{\hat{\eta}}_1 = k_{2sw} \text{sgn}(S_1) \exp(|S_1| - |\varepsilon_1|) \quad (47)$$

where  $k_{2sw}$  is positive parameter and  $k_{2sw} < k_{1sw}$ ,  $\varepsilon_1$  is the threshold of sliding manifold.

**Remark 8.** In initial stage of formation control ( $|S_1| \geq |\varepsilon_1|$ ), equation (47) can be described as  $\dot{\hat{\eta}}_1 = k_{2sw} \exp(|S_1| - |\varepsilon_1|)$ , which means that the greater adaptive rate and faster error convergence will be obtained when the larger item  $|S_1| - |\varepsilon_1|$  is chosen, so it provides an excellent

**Table 1**  
Model parameters of vehicle.

Parameter	$m_{11i}$	$m_{22i}$	$m_{33i}$	$m_{44i}$	
Value	0.00103	0.0150	0.0050	0.00002	
Parameter	$d_{11i}$	$d_{22i}$	$d_{33i}$	$d_{44i}$	$l_{pi}$
Value	0.00042	0.0116	0.002	0.0000075	3.8537

**Table 2**  
Parameters of prescribed boundary.

Parameter	$\bar{e}_{x,0}$	$\underline{e}_{x,0}$	$\bar{e}_{y,0}$	$\underline{e}_{y,0}$	$\bar{e}_{\psi,0}$	$\underline{e}_{\psi,0}$
Value	3	-11	6	-6	8	-5
Parameter	$\bar{e}_{x,t_f}$	$\underline{e}_{x,t_f}$	$\bar{e}_{y,t_f}$	$\underline{e}_{y,t_f}$	$\bar{e}_{\psi,t_f}$	$\underline{e}_{\psi,t_f}$
Value	1.1	-1.1	1.4	-1.4	1	-1
Parameter	$\bar{e}_{\varphi,0}$	$\underline{e}_{\varphi,0}$	$\bar{e}_{\varphi,t_f}$	$\underline{e}_{\varphi,t_f}$	$\bar{k}_j$	$\underline{k}_j$
Value	8	-5	1	-1	-0.5	-0.5

**Table 3**  
Parameters of control.

Parameter	$k_{z_x}$	$k_{z_y}$	$k_{z_\varphi}$	$k_{z_\varphi}$	$k_{1sw}$	$k_{2sw}$
Value	20	0.1	0.8	1	0.5	0.02
Parameter	$k_{3sw}$	$k_{4sw}$	$k_u$	$k_y$	$k_r$	$k_p$
Value	0.5	0.02	0.4	0.1	0.02	0.01
Parameter	$\gamma_u$	$\gamma_h$	$\varepsilon_u$	$\varepsilon_h$		
Value	5	10	0.01	0.01		

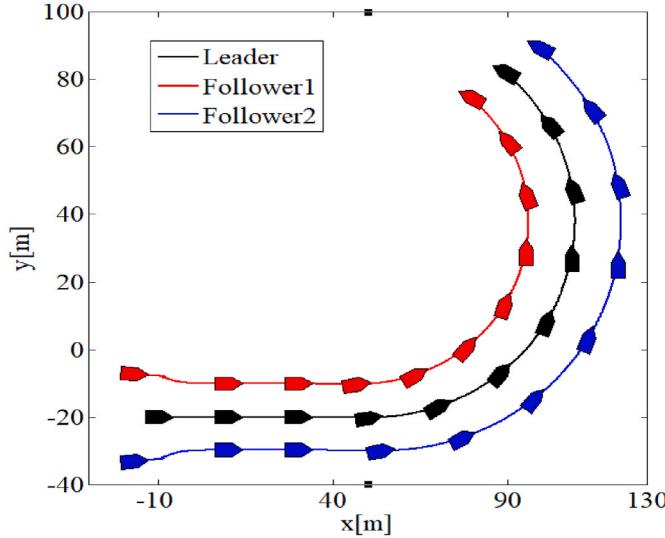


Fig. 5. Leader-follower formation.

transient tracking performance. Once sliding variable satisfies  $|S_1| < |\varepsilon_1|$ ,  $\dot{\hat{\eta}}_1 = -k_{2sw}/\exp(|\varepsilon_1| - |S_1|) < 0$ , which provides higher chattering suppression by reducing switching gain.

**Remark 9.** Sliding variable  $S_1$  will drift away from sliding manifold if switching gain is too small to compensate the lumped uncertainties and approximation errors. To reduce the loss of tracking accuracy caused by the decrease of switching gain, an actor-critic NN-based compensation term  $\tau_{li}^{com}$  is devised to prevent sliding variable from exceeding the designed threshold  $\varepsilon_1$ . Obviously, the compensation term (46) is opposite to the switching gain, which can acquire the reduction of switching gain  $\Delta\hat{\eta}_1$  and returns it to  $\tau_{li}^c$  when  $S_1$  tries to leave the vicinity of sliding manifold.

Based on the analysis of remarks 8–9, the adaptive mechanism (47) and RL compensation term can effectively improve the convergence rate of error when tracking error is large, and appropriately reduce the adaptive rate when tracking error is small, so as to achieve the chattering reduction, address the problem caused by the smaller switching gain and ameliorate formation performance.

Define Lyapunov function candidate  $V_2$  as

$$V_2 = \frac{1}{2}S_1^2 + \frac{1}{2\gamma_{cl}}\tilde{W}_{cl}^T\tilde{W}_{cl} + \frac{1}{2\gamma_u}\tilde{W}_u^T\tilde{W}_u \quad (48)$$

where  $\tilde{W}_u = W_u^* - \widehat{W}_u$ ,  $\tilde{W}_{cl} = W_{cl}^* - \widehat{W}_{cl}$ ,  $\gamma_u, \gamma_{cl}$  are positive parameters.

Differentiating equation (48) along with (40), (42), (45)–(47), we have

$$\begin{aligned} \dot{V}_2 = & -\hat{\eta}_1|S_1| - k_{2sw}|S_1|\exp(|S_1| - |\varepsilon_1|) - k_{1sw}S_1^2 \\ & - \frac{k_{cl}S_1L_1R_1}{\exp(\hat{\eta}_1)} - \frac{k_{cl}S_1W_{cl}^{*T}|L_1R_1|H_{cl}(Z)}{\exp(\hat{\eta}_1)} \\ & + S_1T_u\sigma_u + S_1\mu_1 + \tilde{W}_u^T\left(T_uS_1H_u(Z) - \frac{1}{\gamma_u}\dot{\hat{W}}_u\right) \\ & + \tilde{W}_{cl}^T\left(\frac{k_{cl}S_1|L_1R_1|H_{cl}(Z)}{\exp(\hat{\eta}_1)} - \frac{1}{\gamma_{cl}}\dot{\hat{W}}_{cl}\right) \end{aligned} \quad (49)$$

where  $\hat{\eta}_1$  is given by equation (47) and it is a positive switching gain;  $\sigma_u$  is the NN approximation error and  $|\sigma_u| \leq \bar{\sigma}_u$ , and  $\bar{\sigma}_u$  represents the desired approximation accuracy and  $\bar{\sigma}_u > 0$ .

Design the actor-NN and critic-NN adaptive law  $\dot{\hat{W}}_u, \dot{\hat{W}}_{cl}$  as

$$\begin{aligned} \dot{\hat{W}}_u &= \gamma_u S_1 H_u(Z) (T_u - \hat{\Psi}_{cl}) \\ \dot{\hat{W}}_{cl} &= \gamma_{cl} |L_1 R_1| S_1 H_{cl}(Z) (k_{cl} \exp(-\hat{\eta}_1) - \tilde{W}_u^T H_u(Z)) \end{aligned} \quad (50)$$

**Remark 10.** From equation (50), the updating law of actor-NN  $\dot{\hat{W}}_u$  contains critic function  $\hat{\Psi}_{cl}$ , and critic function is regarded as the reinforcement learning signal and guides the operation of actor-NN to address the lumped uncertainties and optimize control input. Besides, the updating law of critic-NN  $\dot{\hat{W}}_{cl}$  includes actor-NN weight  $\tilde{W}_u^T$ , and critic-NN weight is applied to assess the current formation performance and guides the next update of actor-NN.

Utilizing inequality  $\hat{\eta}_1 > 0, |S_1| > 0, k_{1sw}|S_1|\exp(|S_1| - |\varepsilon_1|) > 0$  and weight update law (50), equation (49) can be rewritten as

$$\dot{V}_2 \leq -k_{1sw}S_1^2 + S_1[\mu_1 + T_u\bar{\sigma}_u + \Phi_1] \quad (51)$$

where  $\Phi_1 = \tilde{W}_u^T H_u(Z)$

$$\begin{aligned} \hat{\Psi}_{cl} &+ \tilde{W}_{cl}^T |L_1 R_1| H_{cl}(Z) \tilde{W}_u^T H_u(Z) - k_{cl} W_{cl}^{*T} |L_1 R_1| H_{cl}(Z) \exp(-\hat{\eta}_1) \\ &- k_{cl} L_1 R_1 \exp(-\hat{\eta}_1). \end{aligned}$$

**Lemma 6.** For surge subsystem, if there exists a positive parameter  $\varphi_1$  and  $\varphi_1 \geq \max[|W_u^{*T} H_u(Z)|, |\tilde{W}_u^T H_u(Z)|, |W_{cl}^{*T} H_{cl}(Z)|]$ , the following inequality holds:

$$\Phi_1 \leq |L_1 R_1| (2\varphi_1 + k_{cl} \exp(-\hat{\eta}_1))(1 + \varphi_1) \quad (52)$$

**Proof.** Based on equation (45),  $\hat{\Psi}_{cl} = L_1 R_1 + |L_1 R_1| \tilde{W}_{cl}^T H_{cl}(Z)$ . The following formula holds

$$\begin{aligned} \tilde{W}_u^T H_u(Z) \hat{\Psi}_{cl} &= (W_u^{*T} - \tilde{W}_u^T) H_u(Z) L_1 R_1 \\ &+ (W_u^{*T} - \tilde{W}_u^T) |L_1 R_1| H_u(Z) \tilde{W}_{cl}^T H_{cl}(Z) \end{aligned} \quad (53)$$

$$\tilde{W}_{cl}^T |L_1 R_1| H_{cl}(Z) \tilde{W}_u^T H_u(Z) = (W_{cl}^{*T} - \tilde{W}_{cl}^T) |L_1 R_1| H_{cl}(Z) (W_u^{*T} - \tilde{W}_u^T) H_u(Z) \quad (54)$$

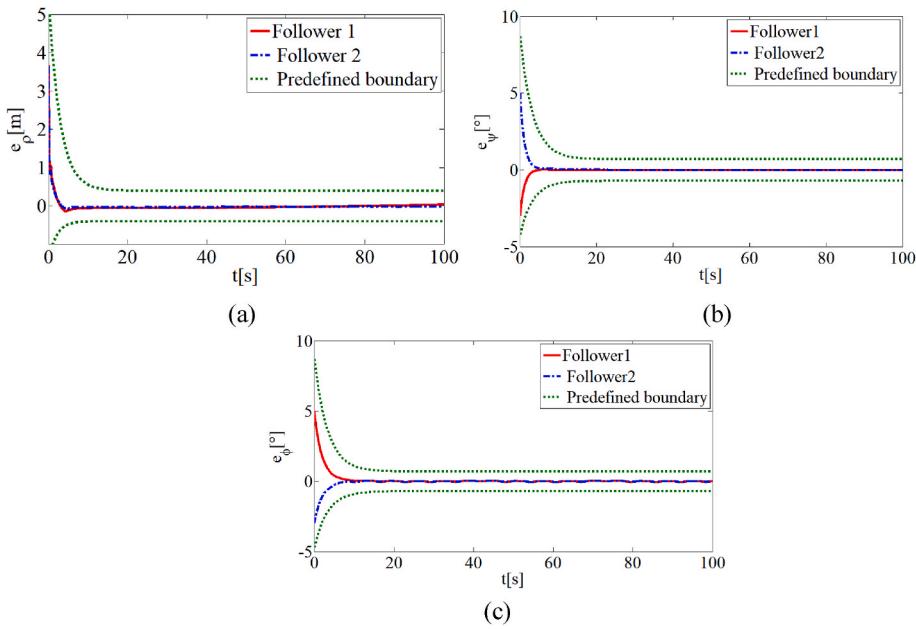
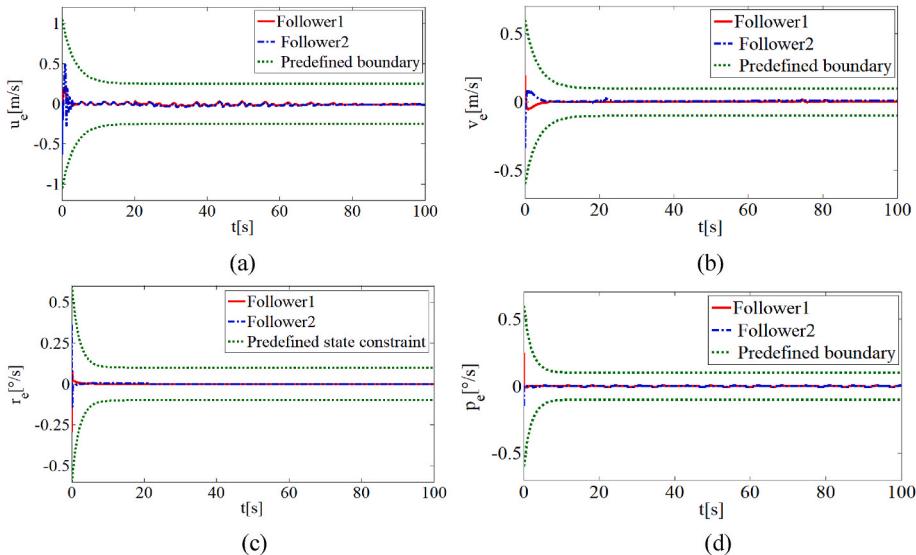
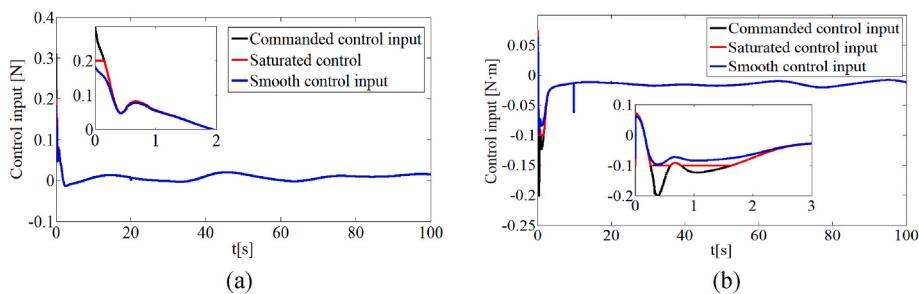
Fig. 6. Formation errors in  $\rho, \psi, \phi$ .Fig. 7. Velocity tracking errors in  $u, v, r, p$ .

Fig. 8. Trajectories of the commanded control input, saturated input, and smooth control input.

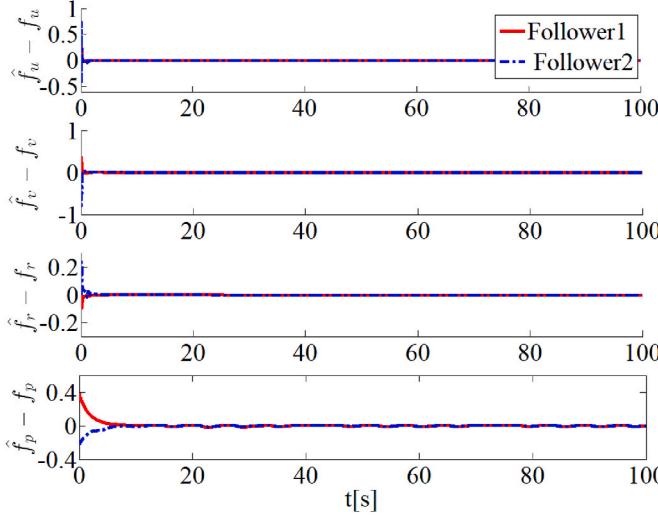


Fig. 9. NN approximate errors of the lumped uncertainties.

Using equations (53) and (54), the following results hold

$$\begin{aligned} & \tilde{W}_u^T H_u(Z) \tilde{\Psi}_{c1} + \tilde{W}_{c1}^T |L_1 R_1| H_{c1}(Z) \tilde{W}_u^T H_u(Z) \\ = & W_u^{*T} H_u(Z) L_1 R_1 - \tilde{W}_u^T H_u(Z) L_1 R_1 + W_u^{*T} H_u(Z) |L_1 R_1| W_{c1}^{*T} H_{c1}(Z) \quad (55) \\ & - \tilde{W}_u^T H_u(Z) |L_1 R_1| W_{c1}^{*T} H_{c1}(Z) \end{aligned}$$

Applying equation (55) and  $\varphi_1 \geq \max[|W_u^{*T} H_u(Z)|, |\tilde{W}_u^T H_u(Z)|, |W_{c1}^{*T} H_{c1}(Z)|]$ , the variable  $\Phi_1$  in equation (51) can be expressed as

$$\begin{aligned} \Phi_1 \leq & |L_1 R_1| [|W_u^{*T} H_u(Z)| + |\tilde{W}_u^T H_u(Z)| + |W_u^{*T} H_u(Z)||W_{c1}^{*T} H_{c1}(Z)|] \\ + & |\tilde{W}_u^T H_u(Z)||W_{c1}^{*T} H_{c1}(Z)| + k_{c1} \exp(-\hat{\eta}_1) + k_{c1} |W_{c1}^{*T} H_{c1}(Z)| \exp(-\hat{\eta}_1) \\ \leq & |L_1 R_1| (2\varphi_1 + k_{c1} \exp(-\hat{\eta}_1)) (1 + \varphi_1) \quad (56) \end{aligned}$$

Based on lemma 6 and the additional control term in (46), equation (51) can be further obtained

$$\dot{V}_2 \leq -k_{1sw} S_1^2 + S_1 [|L_1 R_1| (2\varphi_1 + k_{c1} |S_1| \exp(-\hat{\eta}_1)) (1 + \varphi_1) + \mu_1 + T_u \bar{\sigma}_u] \quad (57)$$

Substituting  $\mu_1$  in equation (46) into (57), we have

$$\dot{V}_2 \leq -k_{1sw} S_1^2 \quad (58)$$

**Step 2** Based on equation (35), DSC is also used in the variables  $\dot{\alpha}_v, \dot{\alpha}_r, \dot{\alpha}_p$ .

$$\varepsilon_h \dot{\beta}_h + \beta_h = \alpha_h, \beta_h(0) = \alpha_h(0), h = v, r, p \quad (59)$$

where  $\varepsilon_h$  is a positive constant,  $\beta_h$  is the filter variable,  $\beta_h(0), \alpha_h(0)$  are initial conditions of  $\beta_h, \alpha_h$ .

According to equation (59), the following equation can be obtained

$$\dot{\beta}_h = -\frac{g_h}{\varepsilon_h} + \dot{\alpha}_h = -\frac{g_h}{\varepsilon_h} + B_h(\cdot) \quad (60)$$

where  $g_h = \beta_h - \alpha_h, B_h(\cdot)$  is the time derivative of  $\alpha_h$ . Based on assumption 2, there exists a positive constant  $B_{hM}$  and  $|B_h(\cdot)| \leq B_{hM}$ .

To satisfy the state constraints and improve the transient performance and robustness in sway-yaw-roll subsystem, the sliding variable based on logarithmic function is defined as

$$S_2 = \sum_{h=v,r,p} \left[ \frac{1}{2} \ln \frac{k_{h_e}^2}{k_{h_e}^2 - (h_e + k_h)^2} - c_2(\beta_h - \alpha_h) \right] + (\varepsilon_{fvp} - \hat{\eta}_2) \quad (61)$$

where  $k_h$  is the design parameter and it helps the control law avoid singularity,  $c_2$  is a positive constant,  $\varepsilon_{fvp}$  is the upper boundary of approximation error,  $\hat{\eta}_2$  is the positive compensation law of approximation error,  $k_{h_e}$  is continuous time-varying functions, which denote the upper boundary of variable  $h_e$ .

Differentiating equation (61) along with (19) and (59), we have

$$\begin{aligned} \dot{S}_2 = & \sum_{h=v,r,p} [T_h(f_{hi} - \dot{\beta}_h) + c_2(B_{hM} - \dot{\beta}_h) - T_{k_h}] \\ & + \frac{T_r}{m_{33i}} (\tau_{2i}^c + \Delta \tau_{2i}^c + d_{ri}) + T_r \xi_2 - \dot{\hat{\eta}}_2 \quad (62) \end{aligned}$$

$$\text{where } T_h = \frac{h_e + k_h}{k_{h_e}^2 - (h_e + k_h)^2}, T_{k_h} = \frac{k_{h_e} (h_e + k_h)^2}{k_{h_e} [k_{h_e}^2 - (h_e + k_h)^2]}.$$

In accordance with the design step of surge controller, actor NN-based approximation equation and critic NN-based critic function are shown as

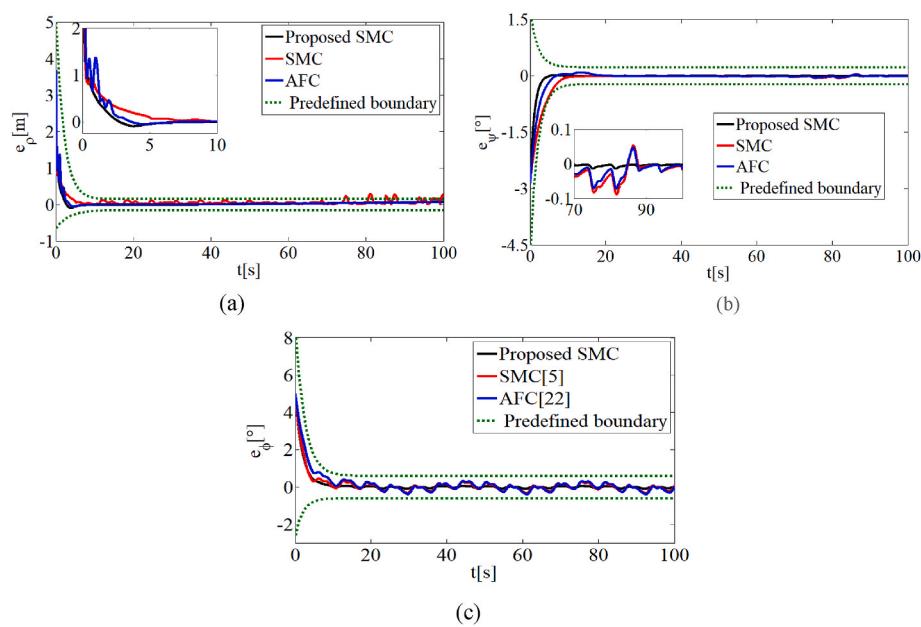


Fig. 10. Formation errors of the vehicle  $i-1$  and follower 1 under different controllers.

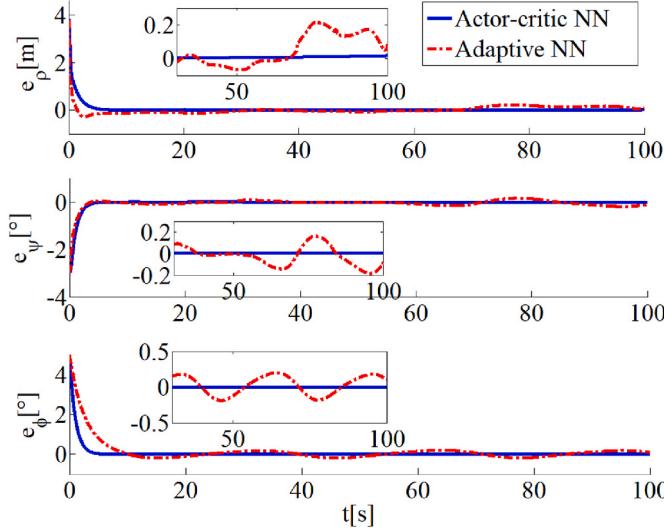


Fig. 11. Formation error compared with different NN architecture.

Table 4  
RMS values of formation errors.

Control Strategies	Range Errors (m)	Orientation Errors (°)	Roll Angle Errors (°)
SMC	0.7724	0.0572	0.3770
AFC	0.2378	0.0536	0.4206
SMC with adaptive NN	0.1052	0.0959	0.1358
Proposed SMC with RL	0.0103	0.0032	0.0063

$$\begin{aligned} f_{hi} &= \widehat{W}_h^T H_h(Z) + \widetilde{W}_h^T H_h(Z) + \sigma_h, h = v, r, p \\ \widehat{\Psi}_{c2} &= L_4 R_2 + |L_4 R_2| \widehat{W}_{c2}^T H_{c2}(Z) \\ R_2 &= \frac{L_5}{1 + \exp(-L_6 S_2)} - \frac{L_5}{1 + \exp(L_6 S_2)} \end{aligned} \quad (63)$$

where  $\widehat{W}_h^T$  is the estimation of  $W_h^{*T}$ ,  $\widetilde{W}_h^T = W_h^{*T} - \widehat{W}_h^T$ ,  $\sigma_h$  is NN approximation error and  $|\sigma_h| \leq \bar{\sigma}_h$ ,  $\bar{\sigma}_h$  denotes a desired precision and  $\bar{\sigma}_h > 0$ ;  $L_5, L_6$  are positive constants,  $R_2 \in [-L_4, L_4]$  increases with the increase of sliding variable  $S_2$  and it corresponds to the performance of formation errors,  $\widehat{W}_{c2}^T$  is the estimation of the ideal weight,  $H_{c2}(Z)$  is the

basis function.

Based on equations (62), (63), the control law  $\tau_{2i}^c$  can be assigned as

$$\begin{aligned} \tau_{2i}^c &= \sum_{h=v,r,p} \frac{m_{33i}}{T_h} [T_h (-\widehat{W}_h^T H_h(Z) + \dot{\beta}_h) - c_2(B_{hM} - \dot{\beta}_h) + T_{kh}] \\ &\quad + \frac{m_{33i}}{T_h} (\tau_{2swi} + \tau_{2i}^{com} + \mu_2) - \Delta \tau_{2i}^c - m_{33i} \dot{\xi}_2 - \widehat{d}_{ri} \\ \tau_{2i}^{com} &= -\frac{k_{c2}}{\exp(\widehat{\eta}_2)} \widehat{\Psi}_{c2} \\ \mu_2 &= -|L_4 R_2| \left[ 2\varphi_2 + \frac{k_{c2}}{\exp(\widehat{\eta}_2)} (1 + \varphi_2) \right] - \sum_{h=v,r,p} T_h \bar{\sigma}_h \end{aligned} \quad (64)$$

where  $\tau_{2i}^{com}$  is a compensation term,  $\mu_2$  is an additional control term,  $\tau_{2swi} = -\widehat{\eta}_2 \text{sgn}(S_2) - k_{3sw} S_2$ ,  $k_{c2}$  is a positive constant,  $\varphi_2$  is a positive design parameter,  $\widehat{d}_{ri}$  is the estimation of  $d_{ri}$  and obtained from lemma 5.

To mitigate chattering in (64), the switching gain adaptive law is chosen as

$$\dot{\widehat{\eta}}_2 = k_{4sw} \text{sgn}(S_2) \exp(|S_2| - |\varepsilon_2|) \quad (65)$$

where  $k_{4sw}$  is positive parameter and  $k_{4sw} < k_{3sw}$ ,  $\varepsilon_2$  is the designed threshold of sliding manifold.

Define the following Lyapunov function candidate as

$$V_3 = \frac{1}{2} S_2^2 + \sum_{h=v,r,p} \frac{1}{2\gamma_h} \widetilde{W}_h^T \widetilde{W}_h + \frac{1}{2\gamma_{c2}} \widetilde{W}_{c2}^T \widetilde{W}_{c2} \quad (66)$$

where  $\widetilde{W}_h = W_h^* - \widehat{W}_h$ ,  $\widetilde{W}_{c2} = W_{c2}^* - \widehat{W}_{c2}$ ,  $\gamma_h, \gamma_{c2}$  are positive parameters

Differentiating equation (66) along with (62)-(65), we have

$$\begin{aligned} \dot{V}_3 &= -\widehat{\eta}_2 |S_2| - k_{3sw} S_2^2 - k_{4sw} |S_2| \exp(|S_2| - |\varepsilon_2|) \\ &\quad + \sum_{h=v,r,p} \widetilde{W}_h^T \left( T_h S_2 H_h(Z) - \frac{1}{\gamma_h} \dot{\widehat{W}}_h \right) - \frac{k_{c2} S_2 L_4 R_2}{\exp(\widehat{\eta}_2)} \\ &\quad - \frac{k_{c2} S_2 W_{c2}^{*T} |L_4 R_2| H_{c2}(Z)}{\exp(\widehat{\eta}_2)} + \mu_2 S_2 + \sum_{h=v,r,p} T_h \sigma_h S_2 \\ &\quad + \widetilde{W}_{c2}^T \left( \frac{k_{c2} S_2 |L_4 R_2| H_{c2}(Z)}{\exp(\widehat{\eta}_2)} - \frac{1}{\gamma_{c2}} \dot{\widehat{W}}_{c2} \right) \end{aligned} \quad (67)$$

Design the actor-NN and critic-NN adaptive law  $\dot{\widehat{W}}_h, \dot{\widehat{W}}_{c2}$  as

$$\begin{aligned} \dot{\widehat{W}}_h &= \gamma_h S_2 H_h(Z) (T_h - \widehat{\Psi}_{c2}), h = v, r, p \\ \dot{\widehat{W}}_{c2} &= \gamma_{c2} S_2 |L_4 R_2| H_{c2}(Z) \left[ k_{c2} \exp(-\widehat{\eta}_2) - \sum_{h=v,r,p} \widehat{W}_h^T H_h(Z) \right] \end{aligned} \quad (68)$$

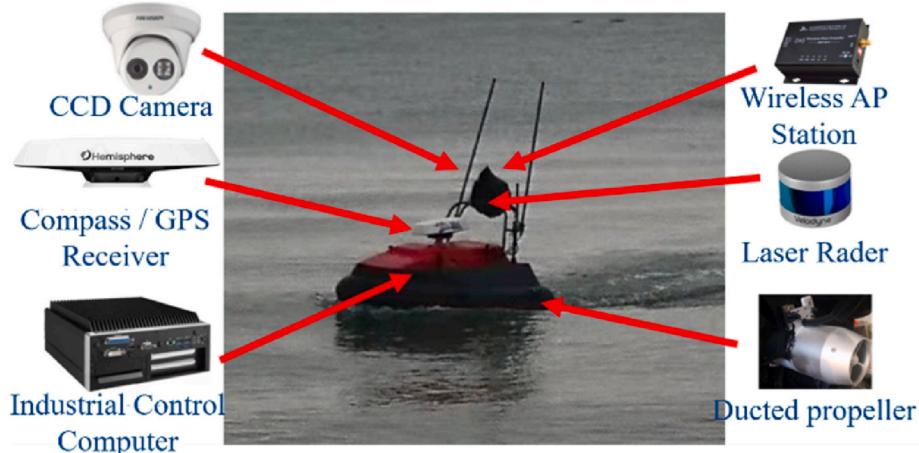


Fig. 12. Experimental platform of USV.

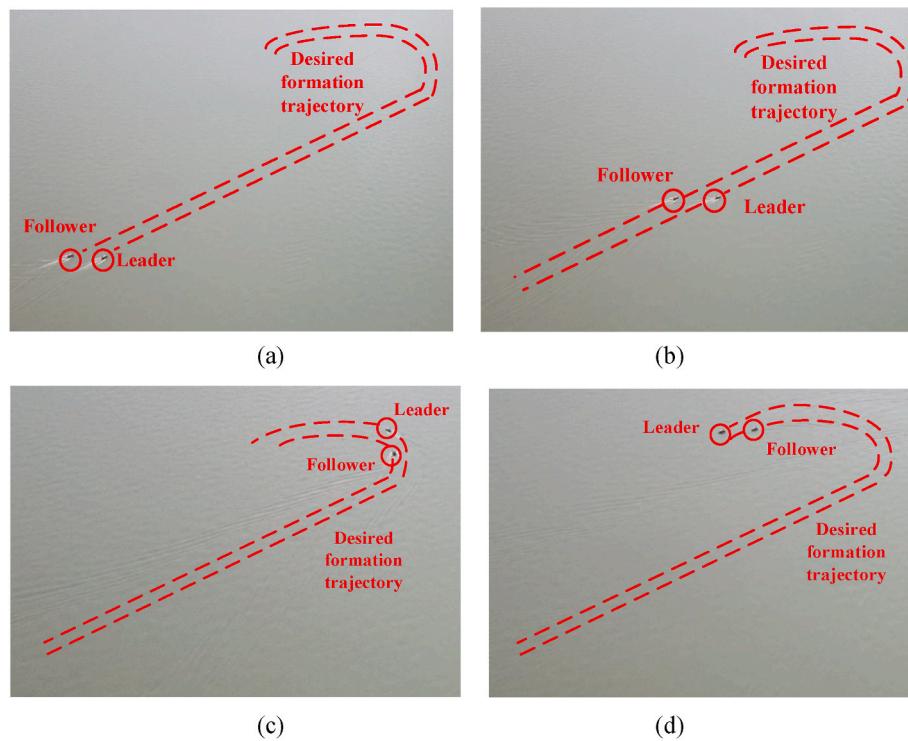


Fig. 13. Leader-follower close formation.

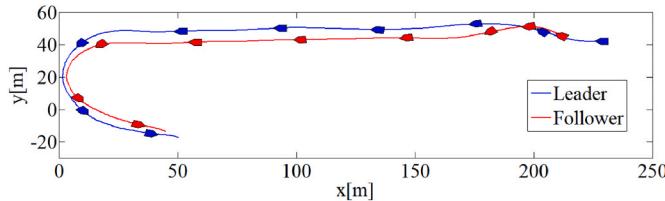


Fig. 14. Leader-follower close formation trajectory.

Using inequality  $\hat{\eta}_2|S_2| > 0$ ,  $\dot{\hat{\eta}}_2 S_2 = k_{2sw}|S_2|\exp(|S_2|-|\varepsilon_2|) > 0$  and weight update law (68), equation (67) can be rewritten as

$$\dot{V}_3 \leq -k_{3sw}S_2^2 + S_2 \left[ \mu_2 + \sum_{h=v,r,p} T_h \bar{\sigma}_h + \Phi_2 \right] \quad (69)$$

where  $\Phi_2 = \sum_{h=v,r,p} \tilde{W}_h^T H_h(Z) \hat{\Psi}_{c2} + \sum_{h=v,r,p} \tilde{W}_h^T H_h(Z) |L_4 R_2| \tilde{W}_{c2}^T H_{c2}(Z) - k_{c2} L_4 R_2 \exp(-\hat{\eta}_2) - k_{c2} W_{c2}^{*T} |L_4 R_2| H_{c2}(Z) \exp(-\hat{\eta}_2)$ .

According to lemma 6, the following inequality holds

$$\begin{aligned} \Phi_2 &\leq |L_4 R_2| \left[ \sum_{h=v,r,p} (|W_h^{*T} H_h(Z)| + |\tilde{W}_h^T H_h(Z)|) + \sum_{h=v,r,p} |W_h^{*T} H_h(Z)| |W_{c2}^{*T} H_{c2}(Z)| \right. \\ &\quad \left. + \sum_{h=v,r,p} |\tilde{W}_h^T H_h(Z)| |W_{c2}^{*T} H_{c2}(Z)| + k_{c2} \exp(-\hat{\eta}_2) + k_{c2} |W_{c2}^{*T} H_{c2}(Z)| \exp(-\hat{\eta}_2) \right] \\ &\leq |L_4 R_2| (2\varphi_2 + k_{c2} \exp(-\hat{\eta}_2)) (1 + \varphi_2) \end{aligned} \quad (70)$$

$$\text{where } \max \left[ \sum_{h=v,r,p} |\tilde{W}_h^T H_h(Z)|, \sum_{h=v,r,p} |W_h^{*T} H_h(Z)|, |W_{c2}^{*T} H_{c2}(Z)| \right] \leq \varphi_2.$$

Based on (70) and the additional control term in (64), equation (69) can be further obtained

$$\begin{aligned} \dot{V}_3 &\leq -k_{3sw}S_2^2 \\ &\quad + S_2 \left[ |L_4 R_2| (2\varphi_2 + k_{c2} |S_2| \exp(-\hat{\eta}_2)) (1 + \varphi_2) + \mu_2 + \sum_{h=v,r,p} T_h \bar{\sigma}_h \right] \end{aligned} \quad (71)$$

Substituting  $\mu_2$  in equation (64) into (71), we have

$$\dot{V}_3 \leq -k_{3sw}S_2^2 \quad (72)$$

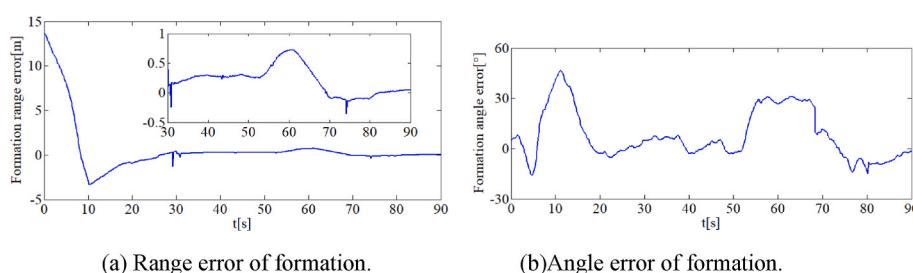


Fig. 15. Errors results of formation.

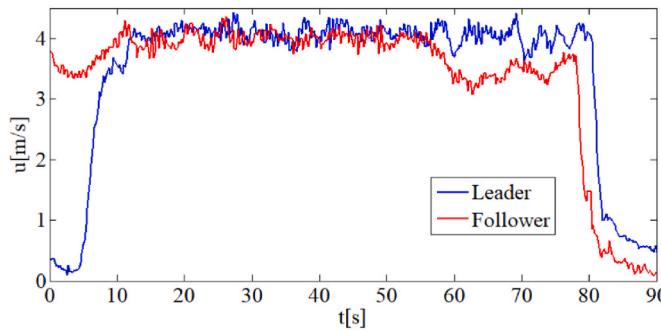


Fig. 16. Velocities of the leader and follower vehicles.

#### 3.4. Finite-time stability analysis

Define the following Lyapunov function candidate as

$$V = V_1 + V_2 + V_3 \quad (73)$$

Utilizing lemma 3, the following inequality can be obtained

$$\begin{aligned} -\left(k_{z_j} z_j^2\right)^l &\leq (1-l)\iota - k_{z_j} z_j^2 \quad (j=x,y,\psi,\varphi) \\ -\left(k_{1sw} S_1^2\right)^l &\leq (1-l)\iota - k_{1sw} S_1^2, \quad -\left(k_{3sw} S_2^2\right)^l \leq (1-l)\iota - k_{3sw} S_2^2 \end{aligned} \quad (74)$$

Differentiating equation (73) along with (57), (72), (74), we have

$$\dot{V} \leq - \sum_{j=x,y,\psi,\varphi} \left(k_{z_j} z_j^2\right)^l - \left(k_{1sw} S_1^2\right)^l - \left(k_{3sw} S_2^2\right)^l + 6(1-l)\iota \quad (75)$$

Equation (75) can be described as

$$\dot{V} \leq - \sum_{m=1}^3 \rho_{V_m} V_m^l + \sum_{m=1}^3 \Delta_m \quad (76)$$

where  $\rho_{V_1} = \min[(2k_{zx})^l, (2k_{zy})^l, (2k_{z\psi})^l, (2k_{z\varphi})^l]$ ,  $\rho_{V_2} = (2k_{1sw})^l$ ,  $\rho_{V_3} = (2k_{3sw})^l$ ,  $\Delta_1 = 4(1-l)\iota$ ,  $\Delta_2 = \Delta_3 = (1-l)\iota$ .

According to equation (76) and lemma 2, the following inequality holds

$$\dot{V} \leq -\rho_V V^l + \Delta \quad (77)$$

where  $\rho_V = \min(\rho_{V_m})$ ,  $\Delta = \sum_{m=1}^3 \Delta_m$ .

Under assumptions 1–4 and lemmas 2–4, consider the close formation control of USV with kinematics (5) and kinetics (19), since  $\rho_V > 0$ ,  $0 < l < 1$ ,  $\Delta = \sum_{m=1}^3 \Delta_m \geq 0$ , the control law (46), (64) and adaptive law (50), (68) guarantee that the prescribed formation performance and time-varying constraints will never be violated and the overall system (73) is finite-time stable.

Based on Wang et al. (2017), the settling time of the finite-time stability can be obtained.

$$T_{reach} = \frac{1}{(1-l)\nu\rho_V} \left[ V^{1-l}(0) - \left( \frac{\Delta}{(1-\nu)\rho_V} \right)^{(1-l)/l} \right] \quad (78)$$

where  $V(0)$  is the initial value of  $V$ , then according to definition 1 and lemma 4, formation errors are SGPFs for  $\forall t > T_{reach}$ .

**Remark 11.** For inequality (76), we obtain  $0 \leq V(t) \leq \Delta/\rho_V + [V(0) - \Delta/\rho_V]e^{-\rho_V t}$ . It is obvious that  $V(t)$  is bounded and exponentially converges to  $\Delta/\rho_V$ . Hence,  $z_x, z_y, z_\psi, z_\varphi, S_1, S_2$  are uniformly bounded according to (73). Since  $\Delta/\rho_V$  can be made arbitrarily small if larger parameter  $k_{zx}, k_{zy}, k_{z\psi}, k_{z\varphi}, k_{1sw}, k_{3sw}, \iota$  and smaller parameter 1 are chosen, the desired formation accuracy can be guaranteed.

#### 4. Simulation validation

In this section, some formation performance and comparison results are included to illustrate excellent performance of the developed controller. The non-dimensional model parameters (Jin et al., 2020) are shown in Table 1.

##### 4.1. Lumped uncertainties

The lumped uncertainties in close formation are mainly composed of the actuator faults, plant uncertainties, external disturbances from ocean and adjacent vehicles, etc. To test actuator faults, the partial loss of effectiveness faults and bias faults are provided as follows:

$$\varpi_{1i} = \begin{cases} 1.0, & t \leq 10 \\ 0.8, & t > 10 \end{cases}, \quad \varpi_{2i} = \begin{cases} 1.0, & t \leq 10 \\ 0.3, & t > 10 \end{cases}, \quad \bar{\tau}_{1i} = 0.01, \quad \bar{\tau}_{2i} = -0.01,$$

The external disturbances are chosen as

$$d_{wi} = [d_{ui}, d_{vi}, d_{ri}, d_{pi}]^T = J^T(\psi)\Phi$$

$$\dot{\Phi} = -K^{-1}\Phi + \Upsilon\Theta$$

where  $\Phi \in R^3$  is intermediate vector,  $J(\psi)$  is a transpose matrix from inertial reference frame to body frame,  $K = \text{diag}(5, 5, 5)$  is a diagonal matrix,  $\Theta \in R^{3 \times 3}$   $\varpi \in R^3$  is the zero-mean Gaussian white noises,  $\Upsilon = \text{diag}(0.05, 0.005, 0.005)$  is a diagonal matrix  $\Upsilon = \text{diag}(0.005, 0.005, 0.005)$ .

To examine the robustness of the RL-based SMC controller, the hydrodynamic damping coefficient and inertia mass coefficient of USV are reduced by 20% in following simulation.

##### 4.2. Formation performance

The close formation consists of three vehicles, i.e. leader  $i-1$ , follower 1 and follower 2. Among them, reference trajectory of vehicle  $i-1$  can be obtained from the following equations

$$\left\{ \begin{array}{l} \dot{x}_{i-1} = u_{i-1} \cos \psi_{i-1} - v_{i-1} \cos \varphi_{i-1} \sin \psi_{i-1} \\ \dot{y}_{i-1} = u_{i-1} \sin \psi_{i-1} + v_{i-1} \cos \varphi_{i-1} \cos \psi_{i-1} \\ \dot{\psi}_{i-1} = r_{i-1} \cos \varphi_{i-1} \\ \dot{\varphi}_{i-1} = p_{i-1} \\ \dot{v}_{i-1} = -\frac{m_{11i-1}}{m_{22i-1}} u_{i-1} r_{i-1} - \frac{d_{22i-1}}{m_{22i-1}} v_{i-1} \end{array} \right.$$

where the yaw angle rate  $r_{i-1}$  of vehicle  $i-1$  satisfies that  $r_{i-1}(t) = 2^\circ/\text{s}$  when  $20\text{s} \leq t \leq 100$  and  $r_{i-1}(t) = 0^\circ/\text{s}$  otherwise.

The desired formation distance and angle are taken as:  $\rho_d = 10 \text{ m}$ ,  $\lambda_{i,des} = 90^\circ$ . Initial states of the leader vehicle  $i-1$ , follower 1 and follower 2 are given as

$$\begin{aligned} [x_{i-1}, y_{i-1}, \psi_{i-1}, \varphi_{i-1}] &= [-10m, -20m, 0^\circ, 0^\circ]^T, \quad [u_{i-1}, v_{i-1}, r_{i-1}, p_{i-1}] \\ &= [2m/\text{s}, 0m/\text{s}, 0^\circ/\text{s}, 0^\circ/\text{s}]^T, \end{aligned}$$

$$\begin{aligned} [x_1, y_1, \psi_1, \varphi_1] &= [-16m, -7.5m, -2.8^\circ, 5^\circ]^T, \quad [x_2, y_2, \psi_2, \varphi_2] \\ &= [-16m, -32.5m, 5^\circ, -3^\circ]^T, \end{aligned}$$

$$[u_1, v_1, r_1, p_1] = [u_2, v_2, r_2, p_2] = [1.5m/\text{s}, 0m/\text{s}, 0^\circ/\text{s}, 0^\circ/\text{s}]^T.$$

In order to guarantee the prescribed performance, function (26) is utilized and some boundary parameters are shown in Table 2. To satisfy the state constraints in (42), the continuous time-varying functions are chosen as  $k_{ue}(t) = k_{ve}(t) = 2e^{-0.5t} + 0.2$ ,  $k_{re}(t) = k_{pe}(t) = 0.5e^{-0.5t} + 0.1$ . In addition, NN architecture consists of 17 nodes with width 2. The ranges of input variable on the surge velocity, sway velocity, yaw

angular rate and roll angular rate are  $[-5 \text{ m/s}, 5 \text{ m/s}]$ ,  $[-3 \text{ m/s}, 3 \text{ m/s}]$ ,  $[-5^\circ/\text{s}, 5^\circ/\text{s}]$ ,  $[-2^\circ/\text{s}, 2^\circ/\text{s}]$ , respectively. Moreover, the control parameters are shown in Table 3.

Fig. 5 displays the formation trajectories of three USVs under lumped uncertainties. Based equations (20) and (22), the formation errors  $e_\rho = \sqrt{e_x^2 + e_y^2} - \rho_d$ ,  $e_\psi$ ,  $e_\varphi$  can be obtained. From Fig. 6 (a)-(c),  $e_\rho$ ,  $e_\psi$ ,  $e_\varphi$  converge to a small neighborhood around zero, which implies that the collision avoidance and rollover constraints will not be violated. Meanwhile, formation errors in Fig. 6 converge with time without overshoot, which conforms to remark 4. As can be seen in Fig. 7 (a)-(d), the velocity errors rapidly converge and meet time-varying state constraints  $k_{u_e}(t)$ ,  $k_{v_e}(t)$ ,  $k_{r_e}(t)$ ,  $k_{p_e}(t)$ . The trajectories of the commanded control input  $\tau_{ki}^c$ , saturated input  $\text{sat}(\tau_{ki}^c)$  and smooth input  $f(\tau_{ki}^c)$  are shown in Fig. 8 (a)-(b). It shows that the sharp point of saturation function at  $|\tau_{ki}^c(t)| = \tau_{KM}$  can be removed by employing smooth input. In Fig. 9, the lumped uncertainties of USVs system can be accurately estimated by the actor-critic NN architecture. From these figures, the results show that the problems of actuator faults, input saturation and input delay are handled and the state responses are smoother with small oscillations. Moreover, the prescribed formation performance together with velocity error constraints can be guaranteed in finite-time.

#### 4.3. Comparison results

From Fig. 10 (a)-(c), the formation tracking results show that the predefined transient and steady-state performances are guaranteed under lumped uncertainties. Compared with SMC (Zhao et al., 2019b) and AFC (Dai et al., 2020), the smaller and smoother formation error can be obtained in a finite-time by the proposed SMC with RL, and initial overshoot can be effectively avoided. Additionally, in order to validate the excellent performance of the proposed scheme with RL algorithm, we present a comparison between adaptive NN-based SMC strategy (Wang et al., 2020a) and the proposed actor-critic NN-based SMC scheme. The same control parameters and external disturbances are selected for the SMC with adaptive NN strategy. The results show that the adaptive NN-based strategy is easy to induce the loss of formation accuracy resulting from the reduced switching gain in Fig. 11. In contrast to the adaptive NN strategy, the trade-off problem among the chattering inherent in SMC, the faster adaptive rate, and the high-accuracy formation performance has been solved effectively by utilizing the proposed actor-critic NN scheme. To further compare the control performance in SMC, AFC, SMC with adaptive NN and the proposed SMC with RL, the root-mean-square (RMS) values of steady-state formation errors (from 30s to 100s) are given in Table 4.

### 5. Experimental results

In this section, the experimental of the proposed control scheme is given to test the correctness of the simulation results and the effectiveness of the proposed control algorithm. After briefly introducing the hardware experimental platform, the experimental results of the close formation are given.

#### 5.1. Experimental platform

The USVs are designed for close formation tasks, as shown in Fig. 12. It is composed of the hull, onboard sensor module, motion control module, communication module, propeller, remote monitoring system. The detailed description of the platform is as follows:

- Hull: The hull is made of carbon fiber and it provides space for the onboard sensors, motion control equipments, communication instruments and propellers.
- Onboard sensor module: The onboard sensors include compass/GPS receiver, laser rader and CCD camera. The compass/GPS receiver

provides the position and attitude of USV. Laser rader is used to detect and obtain information about obstacles. CCD camera is based on color and scale information to assist laser radar for obstacle detection.

- Motion control module: The industrial control computer is applied to generate the PWM signal to control the propeller rate according to the marine tasks.
- Communication module: Wireless AP station is used for data transmission.
- Propeller: USV adopts double ducted propeller, and the yaw moment of USV is controlled by the differential action.
- Remote monitoring system: It is the terminal of human-computer interaction. Using the remote monitoring system, users can monitor the navigation status of USVs, send commands to USVs and receive the position and attitude data from USVs.

#### 5.2. Formation control

In this section, two USVs are given to achieve the close formation, and one is the leader vehicle and the other is the follower vehicle. The trajectory of formation is consistent with Sec. 4.2, which is composed of straight-line and arc. The desired formation range and angle are taken as:  $\rho_d = 9 \text{ m}$ ,  $\lambda_{i,des} = 63^\circ$ . The desired velocity of leader is 4 m/s. Initial states of the leader  $i=1$  and follower 1 are given as

$$[x_{i-1}, y_{i-1}, \psi_{i-1}] = [230m, 43m, 180^\circ]^T, [u_{i-1}, v_{i-1}, r_{i-1}] = [2.5m/s, 0m/s, 0^\circ/s]^T,$$

$$[x_1, y_1, \psi_1] = [212m, 45.8m, 150^\circ]^T, [u_1, v_1, r_1] = [4.1m/s, 0m/s, 0^\circ/s]^T.$$

The close formation results are shown in Fig. 13 (a)-(d). In Fig. 13 (a), USVs receive the formation commands, then the propellers start to work, and the attitudes of USVs are adjusted. In Fig. 13 (b)-(c), the leader and follower vehicles adjust the formation range and angle to the target values to form formation. In Fig. 13 (d), the desired close formation is maintained. In addition, Fig. 14 to Fig. 16 show the trajectory, errors results of formation and velocity, respectively.

Before 50s, USVs were the straight-line formation, and then transformed into the arc formation. Based on Figs. 14 and 15 (a)-(b), the formation gradually stabilized after 30s, and the formation range and angle error converged within 0.5 m and  $10^\circ$  respectively. After 50s, the formation angle error expands. This is because the leader USV tracks the arc trajectory, and then the follower USV receives the update heading data from leader, so as to change the heading and achieve the arc formation. Due to the time-delay of communication, propeller and hull of USVs, the formation angle error lasts about 10s. Fig. 16 shows the velocities of formation. To sum up, the above experimental results have a good effect in close formation control, which verifies the simulation results in Sec. 4 and the effectiveness for the proposed control algorithm.

### 6. Conclusions

The finite-time formation control architecture with prescribed performance and state constraints is introduced for a team of USVs to achieve the leader-follower close formation. A SMC strategy with actor-critic NN is applied to deal with the lumped uncertainties, input saturation and input delay. Then, an exponential boundary function with the terminal time is developed to suppress overshoot more reasonably. Besides, the trade-off between the chattering and adaptive rate has been handled by switching mechanism. Meanwhile, the RL-based compensation term is constructed to handle the formation accuracy problem caused by the reduced switching gain. Numerical simulations and experimental results show the superiority of the developed control method in formation control. Future work is to extend the developed controller to time-delay systems with unknown control coefficients, and improve the controller to realize the scaling, splitting and merging of formation.

## CRediT authorship contribution statement

**Huizi Chen:** Conceptualization, Investigation, Software, Validation, Writing – original draft. **Huaicheng Yan:** Methodology, Writing – review & editing. **Yueying Wang:** Methodology, Writing – review & editing. **Shaorong Xie:** Investigation, Methodology. **Dan Zhang:** Project administration, Supervision.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work is supported by the National Science Fund for Distinguished Young Scholars of China under Grant 61625304, and the National Natural Science Foundation of China under Grant 61973208.

## References

- Awad, T., Elgohary, M., Mohamed, T., 2018. Ship roll damping via direct inverse neural network control system. *Alex. Eng. J.* 57 (4), 2951–2960.
- Baek, S., Baek, J., Han, S., 2019. An adaptive sliding mode control with effective switching gain tuning near the sliding surface. *IEEE Access* 7, 15563–15572.
- Carter, A., 2018. Reflections on American Grand Strategy in Asia. Belfer Center.
- Chan, N.P.K., Jayawardhana, B., de Marina, H.G., 2021. Angle-constrained formation control for circular mobile robots. *IEEE Control Syst. Lett.* 5 (1), 109–114.
- Chen, H., Peng, Y., Wang, Y., Xie, S., Yan, H., 2020. Leader-follower close formation control for underactuated surface vessel via terminal hierarchical sliding mode. *Int. J. Adv. Rob. Syst.* 17 (3), 1–16.
- Dai, S., He, S., Chen, X., Jin, X., 2020. Adaptive leader-follower formation control of nonholonomic mobile robots with prescribed transient and steady-state performance. *IEEE Trans. Ind. Inf.* 16 (6), 3662–3671.
- Dong, Z., Ma, J., Yao, J., 2020. Barrier function-based asymptotic tracking control of uncertain nonlinear systems with multiple states constraints. *IEEE Access* 8, 14917–14927.
- Ellingsen, S.A., 2022. Getting the ducks in a row. *J. Fluid Mech.* 932.
- Guo, Z., Chen, G., 2020. Event-triggered fixed-time cooperative tracking control for uncertain nonlinear second-order multi-agent systems under directed network topology. *J. Franklin Inst.* 357 (6), 3345–3364.
- He, S., Wang, M., Dai, S., Luo, F., 2019. Leader-follower formation control of USVs with prescribed performance and collision avoidance. *IEEE Trans. Ind. Inf.* 15 (1), 572–581.
- Hinostroza, M., Luo, W., Guedes Soare, C., 2015. Robust fin control for ship roll stabilization based on L<sub>2</sub>-gain design. *Ocean Eng.* 94, 126–131.
- Jin, Z., Liu, S., Jin, L., Chen, W., Yang, W., 2020. Model based robust predictive control of ship roll/yaw motions with input constraints. *Appl. Sci.* 10 (10), 3377.
- Lee, J., Chang, P.H., Jin, M., 2017. Adaptive integral sliding mode control with time-delay estimation for robot manipulators. *IEEE Trans. Ind. Electron.* 64 (8), 6796–6804.
- Li, J., Ding, J., Chai, T., Lewis, F., Jagannathan, S., 2022. Adaptive interleaved reinforcement learning: robust stability of affine nonlinear systems with unknown uncertainty. *IEEE Transact. Neural Networks Learn. Syst.* 33 (1), 270–280.
- Li, S., Wang, Q., Ding, L., An, X., Deng, Z., 2020a. Adaptive NN-based finite-time tracking control for wheeled mobile robots with time-varying full state constraints. *Neurocomputing* 403, 421–430.
- Li, Y., Wang, C., Liang, D., Li, X., 2020b. Distributed adaptive leader-following consensus control of connected Lagrangian systems with input hysteresis quantization and time-varying control gain. *J. Franklin Inst.* 357 (2), 942–977.
- Liang, H., Zhang, Y., Huang, T., Ma, H., 2020. Prescribed performance cooperative control for multiagent systems with input quantization. *IEEE Trans. Cybern.* 50 (5), 1810–1819.
- Liang, L., Wen, Y., 2019. Rudder roll stabilization with disturbance compensation model predictive control. *J. Mar. Sci. Technol.* 24, 249–259.
- Liu, C., 2019. Neural-network-based distributed formation tracking control of marine vessels with heterogeneous hydrodynamics. *IEEE Access* 7, 150141–150149.
- Liu, C., Zou, Z., Li, T., 2015. Path following of underactuated surface vessels with fin roll reduction based on neural network and hierarchical sliding mode technique. *Neural Comput. Appl.* 26 (7), 1525–1535.
- Liu, F., Hua, Y., Dong, X., Li, Q., Ren, Z., 2019. Adaptive fault-tolerant time-varying formation tracking for multi-agent systems under actuator failure and input saturation. *ISA Trans.* 1–9.
- Liu, Y., Zhu, Q., Wen, G., 2022. Adaptive tracking control for perturbed strict-feedback nonlinear systems based on optimized backstepping technique. *IEEE Transact. Neural Networks Learn. Syst.* 33 (2), 853–865.
- Liu, Z., Li, G., Zhang, J., Zheng, L., 2020. Path following for underactuated surface vessels with disturbance compensating predictive control. *Int. J. Adv. Rob. Syst.* 17 (2), 1–11.
- Mobayen, S., 2016. Design of lmi-based sliding mode controller with an exponential policy for a class of underactuated systems. *Complexity* 21 (5), 117–124.
- Mobayen, S., Tchier, F., 2017. Nonsingular fast terminal sliding mode stabilizer for a class of uncertain nonlinear systems based on disturbance observer. *Sci. Iran.* 24 (3), 1410–1418.
- Ouyang, Y., Dong, L., Wei, Y., Sun, C., 2020. Neural network based tracking control for an elastic joint robot with input constraint via actor-critic design. *Neurocomputing* 409, 286–295.
- Wang, F., Chen, B., Lin, C., Zhang, J., Meng, X., 2017. Adaptive neural network finite-time output feedback control of quantized nonlinear systems. *IEEE Trans. Cybern.* 48 (6), 1839–1848.
- Wang, J., Wang, C., Wei, Y., Zhang, C., 2020a. Sliding mode based neural adaptive formation control of underactuated auvs with leader-follower strategy. *Appl. Ocean Res.* 94, 101971.
- Wang, N., Gao, Y., Zhang, X., 2021. Data-driven performance-prescribed reinforcement learning control of an unmanned surface vehicle. *IEEE Transact. Neural Networks Learn. Syst.* 32 (12), 5456–5467.
- Wang, N., Gao, Y., Zhao, H., Ahn, C.K., 2020b. Reinforcement learning-based optimal tracking control of an unknown unmanned surface vehicle. *IEEE Transact. Neural Networks Learn. Syst.* 32 (7), 3034–3045.
- Wang, N., Sun, Z., Yin, J., Su, S., Sharma, S., 2018. Finite-time observer based guidance and control of underactuated surface vehicles with unknown sideslip angles and disturbances. *IEEE Access* 6, 14059–14070.
- Wang, Y., Jiang, X., She, W., Ding, F., 2019. Tracking control with input saturation and full-state constraints for surface vessels. *IEEE Access* 144741–144755.
- Xie, Z., Sun, T., Kwan, T., Mu, Z., Wu, X., 2020. A new reinforcement learning based adaptive sliding mode control scheme for free-floating space robotic manipulator. *IEEE Access* 8 (99), 127048–127064.
- Yue, C., Wang, F., Cao, X., Shen, Q., Chen, X., 2020. Robust fault-tolerant attitude tracking with guaranteed prescribed performance. *J. Franklin Inst.* 357 (1), 229–253.
- Zhang, R., Li, J., Jiao, J., 2020. Adaptive fuzzy control for nontriangular form systems with time-varying full-state constraints. *Int. J. Adapt. Control Signal Process.* 34, 919–936.
- Zhao, J., Liang, C., Zhang, X., 2020a. Rudder roll stabilization based on arc tangent nonlinear feedback for ships. *J. Mar. Sci. Eng.* 8 (4), 1–13.
- Zhao, P., Liang, L., Zhang, S., Ji, M., Yuan, J., 2019a. Simulation analysis of rudder roll stabilization during ship turning motion. *Ocean Eng.* 189, 106322.1–106322.11.
- Zhao, S., Wang, X., Chen, H., Wang, Y., 2020b. Cooperative path following control of fixed-wing unmanned aerial vehicles with collision avoidance. *J. Intell. Rob. Syst.* 100 (3–4), 1569–1581.
- Zhao, W., Liu, H., Lewis, F.L., 2020c. Robust formation control for cooperative underactuated quadrotors via reinforcement learning. *IEEE Transact. Neural Networks Learn. Syst.* 32 (10), 4577–4587.
- Zhao, Y., Zhang, Y., Lee, J., 2019b. Lyapunov and sliding mode based leader-follower formation control for multiple mobile robots with an augmented distance-angle strategy. *Int. J. Control. Autom. Syst.* 17 (5), 1314–1321.
- Zheng, Z., Ruan, L., Zhu, M., Guo, X., 2020. Reinforcement learning control for underactuated surface vessel with output error constraints and uncertainties. *Neurocomputing* 399, 479–490.