Flow control of three-dimensional cylinders transitioning to turbulence via multi-agent reinforcement learning

Suárez, Pol; Alcántara-Ávila, Francisco; Rabault, Jean; Miró, Arnau; Font, Bernat; Lehmkuhl, Oriol; Vinuesa, Ricardo

DOI
10.1038/s44172-025-00446-x

Publication date
2025

Document Version
Final published version

Published in
Communications Engineering

Citation (APA)
Suárez, P., Alcántara-Ávila, F., Rabault, J., Miró, A., Font, B., Lehmkuhl, O., & Vinuesa, R. (2025). Flow control of three-dimensional cylinders transitioning to turbulence via multi-agent reinforcement learning. *Communications Engineering*, *4*(1), Article 113. https://doi.org/10.1038/s44172-025-00446-x

communications engineering

A Nature Portfolio journal

Article

https://doi.org/10.1038/s44172-025-00446-x

Flow control of three-dimensional cylinders transitioning to turbulence via multi-agent reinforcement learning

Pol Suárez [1] ✉, Francisco Alcántara-Ávila [1], Jean Rabault [2], Arnau Miró [3], Bernat Font [4], Oriol Lehmkuhl[3] & Ricardo Vinuesa [1] ✉

Active flow control strategies for three-dimensional bluff bodies are challenging to design, yet critical for industrial applications. Here we explore the potential of discovering novel drag-reduction strategies using deep reinforcement learning. We introduce a high-dimensional active flow control setup on a three-dimensional cylinder at Reynolds numbers (Re_D) from 100 to 400, spanning the transition to three-dimensional wake instabilities. The setup involves multiple zero-net-mass-flux jets and couples a computational fluid dynamics solver with a numerical multi-agent reinforcement learning framework based on the proximal policy optimization algorithm. Our results demonstrate up to 16% drag reduction at $Re_D = 400$, outperforming classical periodic control strategies. A proper orthogonal decomposition analysis reveals that the control leads to a stabilized wake structure with an elongated recirculation bubble. These findings represent the first demonstration of training on three-dimensional cylinders and pave the way toward active flow control of complex turbulent flows.

The transportation industry, and the aerospace sector in particular, requires new ground-breaking methods to overcome the challenges that they are currently facing, i.e., the need to reduce fossil fuel-related emissions. The implementation of flow-control systems, both passive and active, plays a vital role in the development of more sustainable solutions that can drastically reduce fuel usage, mitigate air and noise pollution, and even improve maneuverability[1]. The aerodynamic drag in subsonic aircraft is divided mainly into pressure, skin friction (due to viscous stresses), and lift-induced components. Wing-tip effects aside, the two former are the dominant terms.

Control devices utilize aerodynamic principles to manipulate pressure and viscosity, effectively reducing drag. For instance, slats and flaps are control surfaces located on the leading and trailing edges of an airfoil, which impact the aircraft operational conditions[2]. Modern advancements include winglets[3] aimed at mitigating the lift-induced drag or vortex generators[4], which are used to control boundary-layer separation. Such developments have significantly improved aerodynamic performance. Additionally, a number of alternative approaches like morphing surfaces, spiroids, or blowing devices[5,6] are currently being explored. Despite their extensive potential, designing optimal geometries or strategies for these devices has raised significant challenges due to the massive computational resources required to tackle such an intricate interplay between pressure and viscous effects in all flight regimes. Nevertheless, ongoing research efforts and current computational innovations are leading to further advancements toward achieving optimal global control.

In parallel with the recent innovations in flow control, the irruption of machine-learning (ML) techniques has brought tremendous potential to the aeronautics industry, both in terms of studying fundamental problems in fluid mechanics[7,8] and devising completely new strategies for active and passive flow control (AFC and PFC, respectively)[9]. Deep reinforcement learning (DRL) is one of the fastest-growing fields within ML[10] and one of the techniques attracting the most interest. Expanding on its success in tabletop games[11], DRL works well in any system where a controller interacts with an environment to improve a task. That is the case for most AFC cases since DRL interacts with the flow on the fly and receives feedback from it, gaining experience and progressively improving the choice of actions.

AFC setups are complex, high-dimensional problems that require substantial computational resources to find the optimal values within the large parametric space of the control system. DRL and neural networks have emerged as valuable tools to make this process feasible, enabling the development of effective control strategies at a reasonable computational cost. The literature on DRL for AFC grows at a fast pace, exhibiting studies on flow control for two-dimensional (2D) cylinders ranging from $Re_D = 100$ and 2000 (where Re_D is the Reynolds number based on inflow velocity U_∞ and cylinder diameter D) with 17% and 38% drag reduction,

[1]FLOW, Engineering Mechanics, KTH Royal Institute of Technology, Stockholm, Sweden. [2]Independent researcher, Oslo, Norway. [3]Barcelona Supercomputing Center, Barcelona, Spain. [4]Faculty of Mechanical Engineering, Technische Universiteit Delft, Delft, The Netherlands. ✉e-mail: polsm@kth.se; rvinuesa@mech.kth.se

respectively[12–18], aircraft wings[19], fluid-structure interaction[20], turbulent channels[21], shape optimization[22–24], Rayleigh–Bénard convection[25] or turbulence modeling[26–29]. Some recent literature demonstrates the possibility of transfer learning from exploration done in 2D cylinders to 3D domains and higher Re_D: in ref. 30, the wake of a cylinder is controlled by means of two rotating cylinders and in ref. 18, the control is carried out through multiple jets over the cylinder surface. The present work extends this state-of-the-art in 3D cylinders, considering multiple actuators governed by the novel implementation of a multi-agent reinforcement learning (MARL) framework into a setup based on a distributed-input distributed-output (DIDO) scheme. In our case, the agent focuses on exploring the underlying 3D physics during training, and the AFC is implemented by multiple independent zero-net-mass-flow jets placed along the cylinder span and aligned along two slots on the top and bottom surfaces. To the best of the authors' knowledge, this work marks the first time that exploration sessions are directly conducted within 3D cylinders.

As the Reynolds number increases, the flow around a cylinder exhibits different characteristics. Initially, up to approximately $Re_D \approx 40$, steady laminar flow prevails, characterized by symmetric counter-rotating vortices in the near-wake. Beyond $Re_D \approx 190$, laminar vortex shedding emerges, forming the well-known Kármán vortex street. In the subsequent regimes, between $190 < Re_D < 260$, the mode-A instability, characterized by dominant spanwise wavelengths of $\lambda_z/D = 4$[31,32] is dominant. As $Re_D \approx 260$ is surpassed, mode B becomes predominant, and finer three-dimensional features with shorter wavelengths of $\lambda_z/D = 1$ are found. Beyond these regimes, the cylinder wake evolves into a more chaotic and turbulent state.

Discovering flow-control strategies for the flow around a cylinder when the wake transitions from 2D to 3D is challenging. The MARL setup needs to exploit the characteristics of the spanwise structures as the wake becomes three-dimensional to devise effective control approaches. The transition range for 2D environments has been shown to be suitable, showcasing the generalization ability of deep neural networks[12]. However, it has been widely recognized that studying $Re_D > 250$ in a 2D context leads to inaccurate predictions of aerodynamic forces. In the present work, the exploration of the 3D context allows tackling possible novel strategies that take advantage of the drag reduction originated by 3D instabilities.

DRL is based on maximizing a reward r_t provided to an agent interacting continuously with an environment through actions a_t. The agent receives information about the environment state at each actuation step through partial observations s_t of the system. This way, the agent works on the optimization of a policy $\pi(a_t|s_t)$. A sequence of consecutive actions is denoted as an episode. When a batch M of episodes is finished, the agent updates the neural-network weights to progressively determine a policy that maximizes the expected reward for a given s_t. For a detailed understanding of the most recent advances in flow control with MARL, we refer to refs. 33,34.

Results

This study presents our findings on high-dimensional distributed forcing using multiple jets aligned along the spanwise direction of a 3D infinite cylinder. Training was carried out using MARL, which demonstrates superior performance compared to conventional single-agent reinforcement learning (SARL) methods. In the subsequent sections, we explore the development of training strategies $\pi(a_t|s_t)$ aimed at achieving high rewards. We then evaluate the optimal model and compare it with the uncontrolled cases and the periodic control (PC). Note that the uncontrolled converged results across Re_D are obtained after a grid-independence study-more details can be found in Supplementary Documentation Section 1. For the study of the periodic control, more details are also provided in Supplementary Documentation Section 2, where the tables and heatmaps from the optimization problem are presented. These results will be used for the subsequent analysis. In doing so, we investigate the utilization of the trained DRL agents for exploitation without involving any exploration. Statistical analyses are conducted to elucidate the underlying physical mechanisms responsible for drag reduction, leading to potential energy savings.

Training

To ensure an effective DRL training process, it is crucial to precisely define rewards, penalties, action ranges, and a representative environment state. A fundamental aspect of training is leveraging the physical understanding of the controlled phenomenon to evaluate anticipated reward values and physical control strategies thoughtfully. With these considerations in mind, Fig. 1a shows the training curves for the four investigated cases in this study, at Reynolds numbers $Re_D = 100$, 200, 300, and 400. Commonly, sequences of actions a_t, states s_t, and rewards r_t are referred to as "environment episodes". However, in this case, it is more appropriate to call them "pseudo-environment episodes" due to the difference between SARL and MARL, where MARL involves multiple pseudo-environments per environment – $n_{\text{pseudoenvs}} = n_{\text{jets}}$ in this case. Hence, Fig. 1a shows all the final rewards from the raw pseudo-environments, together with the pure drag reduction and lift-biased penalization contributions (see the Methods section for more details). As an example, the $Re_D = 300$ scenario closely resembles the ideal training condition. This is because the curves exhibit minimal lift bias and result in a total reward that matches the pure drag reduction, stabilizing at R values that are manageable and simple to track. Similar patterns are obtained for the other Re_D cases, indicating that the discovered policies are promising for all the cases. Note that we also observe several instances of apparent unlearning, such as for $Re_D = 400$ at episode 500 approximately. This is due to additional exploration of the agent, aimed at increasing the lift asymmetry (note the decrease of the blue line), but quickly returning to exploiting what the agent has identified as a well-performing policy. Once the reward value settles around a certain converged value of $R \approx 1.0$, the training process is concluded. The moment of stopping the training is always associated with a high risk of both unlearning and having a much greater computational burden than expected. The best way to assess when training is concluded is to test the model with no exploration at different stages and observe any drop or improvement in performance. Then, we proceed to the assessment of the learned policies in deterministic mode, which is discussed next.

In terms of computational cost, training is the most significant part. On average, each training session requires about 1200 MARL episodes, which is equivalent to running 120 numerical simulations for the entire domain. All exploration sessions were conducted on the Dardel high-performance computer in the PDC Center at KTH Royal Institute of Technology. The sessions run on 8 nodes simultaneously, each running one numerical simulation comprising 10 simultaneous pseudo-environments. Hence, 80 pseudo-environments in total. Each node has two AMD EPYC™ Zen2 2.25 GHz 64-core processors with 512 GB of memory. With each batch of 8 simulations taking ideally five hours in this particular architecture, it requires less than four days of continuous operation. Moreover, as the Reynolds number (Re_D) increases, the computational cost increases significantly.

Exploitation of the models

At this point, the agent policies are evaluated without any exploration. As a result, the agent calculates the most likely value of the action a_t within its learned probability distribution, aiming to maximize the expected reward. In Fig. 1b, we show the temporal evolution of the DRL control and the PC. The DRL-based control exhibits a clear two-phase process that starts with a short transient period followed by the stationary control policy. We included the transition phase between $t = 0$ to around 20 convective time units. Note that the vortex-shedding period is $T_k = 1/St \approx 1/0.2 = 5$, where $St = fD/U_\infty$ is the Strouhal number and f is the vortex-shedding frequency. These results imply that it takes less than $4T_k$ to reach the stationary behavior. The DRL-based control exhibits a first suction/ejection overshoot, which destabilizes the wake, and then it proceeds to re-stabilize it in a second phase. During the latter, the jet mass flux exhibits lower values, which barely reach 75% of those in the transient overshoots.

The time window shown in Fig. 1b is limited to 125 time units to better highlight the two-phase behavior of the DRL control. However, the actual simulations were run for at least 200 time units to ensure temporal

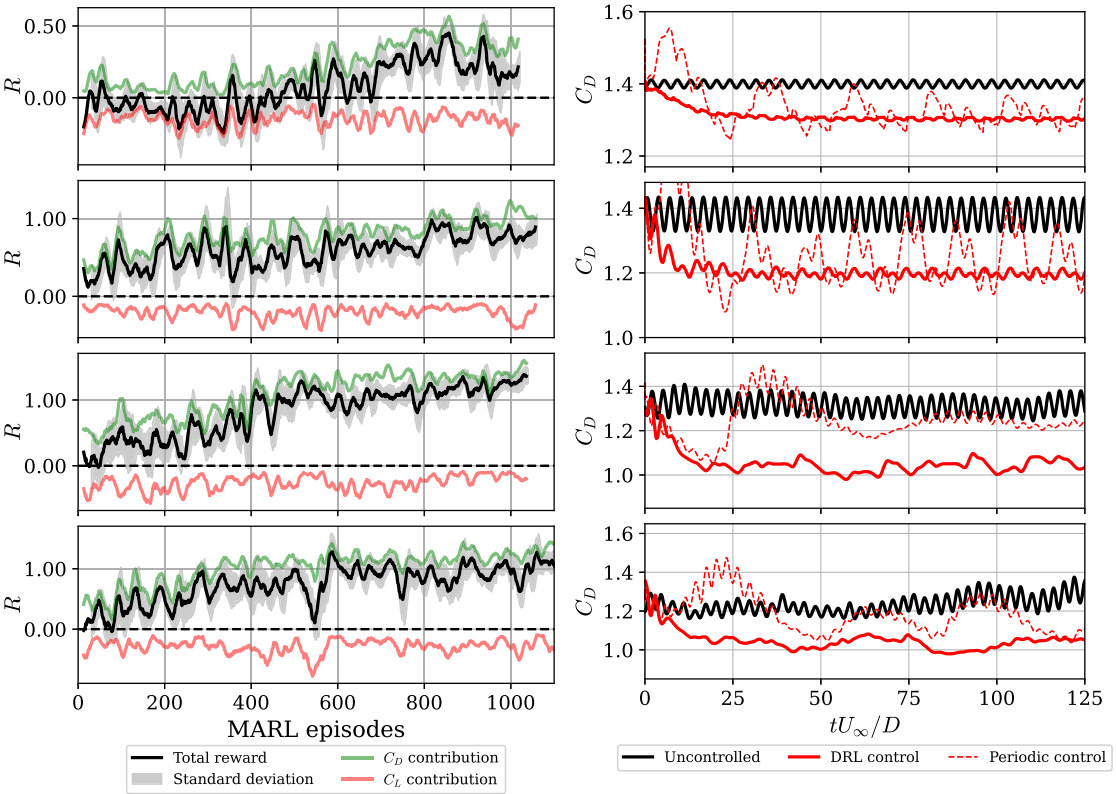

Fig. 1 | Reward evolution during exploration episodes and exploitation of the policies. a Final total reward R, along with its lift-bias and pure drag-reduction components, evaluated during exploration across pseudo-environments or MARL episodes within training sessions. Signals are smoothed by a moving average of 15 values, and the gray shaded area corresponds to the minimum and maximum rewards over those 15 episodes. **b** Drag-coefficient evolution during exploitation of the model. Comparison between uncontrolled, DRL control, and periodic control (PC). From top to bottom, $Re_D = 100$, 200, 300, and 400. The time is non-dimensionalized as tU_∞/D, where t is the physical time, U_∞ the freestream velocity, and D the characteristic length, the cylinder diameter.

convergence, with statistical quantities computed over the entire duration, excluding the initial transients.

This control strategy persists until control stabilizes into stationary behavior, which is monitored by assessing mean quantities and fluctuations in aerodynamic forces. The averaged drag-reduction results for all Re_D are reported in Fig. 2a. It is important to note that all the cases lead to effective drag-reduction rates. The overall performance is much better than what can be obtained with the classical PC strategies. In summary, the percentage changes in the mean drag coefficient are $\Delta\overline{C_D}|_{\mathrm{DRL}} = -7\%, -13.4\%, -21.2\%, -16.2\%$, compared to $\Delta\overline{C_D}|_{\mathrm{PC}} = -5.7\%, -9.7\%, -5.9\%, -5.9\%$ for $Re_D = 100$, 200, 300, and 400, respectively. We acknowledge that there is also a notable reduction in lift oscillations as a consequence of the drag reduction mechanisms employed by the DRL agents, which were not directly addressed in the reward r_t function. This trade-off is an important aspect and presents an opportunity for future approaches that may consider maximizing lift RMS or other metrics depending on the application. All quantities of interest are averaged in time by considering an interval of at least $20T_k$, i.e. over 100 time units, excluding the transients obtained after applying the control. The root-mean-square of the fluctuations, $\phi_{\mathrm{RMS}} = \sqrt{(1/n)\sum_{i=1}^{n}(\phi_i - \overline{\phi})^2}$, minimum and maximum values provide deeper insights into the mentioned robustness. While the mean values alone may suggest a good performance of the PC, the merits of the control should not be assessed solely based on this quantity. When considering an optimal control strategy, the preferred choice typically involves selecting a control with minimal variability and few extreme values, which are characteristics exhibited as evidenced by the DRL-based control.

We also study the ratio between the total fluid mass intercepted by the frontal area of the cylinder E_∞, where $Q_\infty = DU_\infty$, and the total mass used by the actuators E_c. Based on the definitions used in ref. 18, we propose the following expression for the ratio E_c^*:

$$E_c^* = \frac{E_c}{E_\infty} = \frac{L_{\mathrm{jet}}}{(t_2 - t_1)Q_\infty L_z}\int_{t_1}^{t_2}\sum_{i=1}^{n_{\mathrm{jets}}}|Q_i(t)|\,\mathrm{d}t, \tag{1}$$

where t_1 and t_2 define the start and end of our time interval for evaluating the control. Note that while this evaluation of mass consumption is based on a numerical approach with modified Dirichlet boundary conditions, the actual mass cost would depend on the specifics of the experimental setup, including the type of actuators, such as membranes or reservoirs, and the jet configurations used. In the absence of experimental data for these components, we adopt this numerical approach. However, the present results indicate that the mass cost associated with DRL-based control remains minimal in comparison to the significant drag reduction achieved. This emphasizes the efficiency of DRL-based control strategies, as illustrated by the results in Fig. 2, where the $E_c^*/\Delta\overline{C_D}$ ratio demonstrates that DRL requires only half the mass of classical control methods to achieve a comparable drag reduction.

We provide additional physical insight by assessing the power-spectral density (PSD) of the streamwise velocity, shown in Fig. 2b. This figure illustrates how the change of frequency impacts the wake topology after

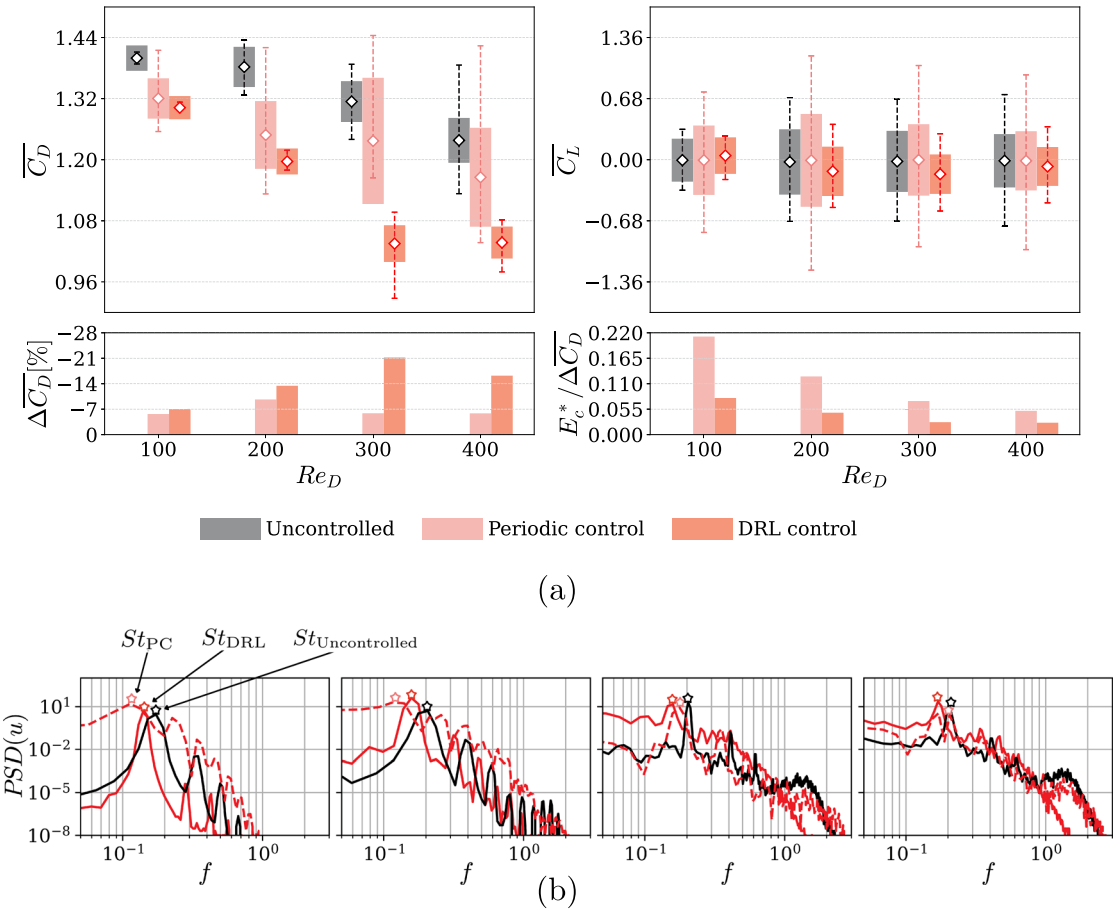

Fig. 2 | Summary of aerodynamic forces and wake spectra for the various studied cases. a Mean drag (C_D) and lift (C_L) coefficients (white diamonds), RMS fluctuations (thick bars), and max-min values range (dashed intervals). Percentage drag reduction $\Delta\overline{C_D}$ and cost metric $E_c^*/\Delta\overline{C_D}$ (lower is better) from Equation (1). **b** Power-spectral density of streamwise velocity u at $x/D = 10.5$ for uncontrolled (black), PC (red dashed), and DRL-based control (red). The dominant frequencies are represented in terms of their Strouhal numbers: St_{PC} for periodic control, St_{DRL} for DRL control, and $St_{Uncontrolled}$ for the natural shedding frequency in the uncontrolled case, where $St = fD/U_\infty$, with f the shedding frequency, D the characteristic length, and U_∞ the freestream velocity. From left to right: $Re_D = 100$, 200, 300, and 400.

Table 1 | Summary of control strategies statistics

	Williamson[31]	Uncontrolled	Periodic control				DRL control			
Re_D	St		Q_{max}	Q_{RMS}	f_c	St	Q_{max}	Q_{RMS}	f_c	St
100	0.165	0.170	0.053	0.037	0.115	0.113	0.016	8.4×10^{-3}	0.144	0.139
200	0.185 (Mode A)	0.186	0.053	0.037	0.130	0.117	0.031	9.5×10^{-3}	0.160	0.157
300	0.203 (Mode B)	0.206	0.018	0.013	0.175	0.177	0.031	7.6×10^{-3}	0.158	0.159
400	0.205 (Mode B)	0.202	0.012	0.008	0.172	0.194	0.025	5.7×10^{-3}	0.171	0.171

Strouhal number St (compared with literature), max and RMS mass-flow rates (Q_{max}, Q_{RMS}), and control frequency f_c (the optimum found for the sinusoidal signal of PC, and the peak from DRL actuators signal).

applying the various control strategies. In particular, both the DRL-based control and PC cases exhibit a reduction in St. Further insight into the various control strategies is provided in Table 1, where several characteristic variables of the various controls are shown. The first important observation is the fact that the root-mean-square (RMS) of the jet mass-flow rate (computed by averaging in time and the spanwise direction) is one order of magnitude lower in the DRL than in the PC. This indicates that the DRL-based control strategies lead to more stable and robust configurations, avoiding large peak-to-peak variations in the actuation. Also, note that we show the control frequencies f_c for the PC (where we report the optimal control frequency) and the DRL (where we report the dominant frequency). Although the f_c values are not dramatically different in the PC and DRL cases, the latter exhibit more complex control laws than the former. It is worth noting that we conducted a parametric study to assess the optimal values of f_c and Q for classical PC, as discussed in the Methods section and referred to the Supplementary Documentation Section 2.

Figure 3 demonstrates the main advantage of a MARL implementation: the control policy can act locally, exploiting wake vortical structures and distributing the jet flow in the spanwise direction to minimize overall drag. In Fig. 3a, we show the temporal and spanwise evolution of the mass-flow rate per unit length from the jets under the DRL-based control strategy. The agent utilizes less than 10% of the maximum possible value, as discussed in the Methods section. For $Re_D \leq 200$, the mass flow is uniformly distributed in the spanwise direction, while beyond this Reynolds number, the

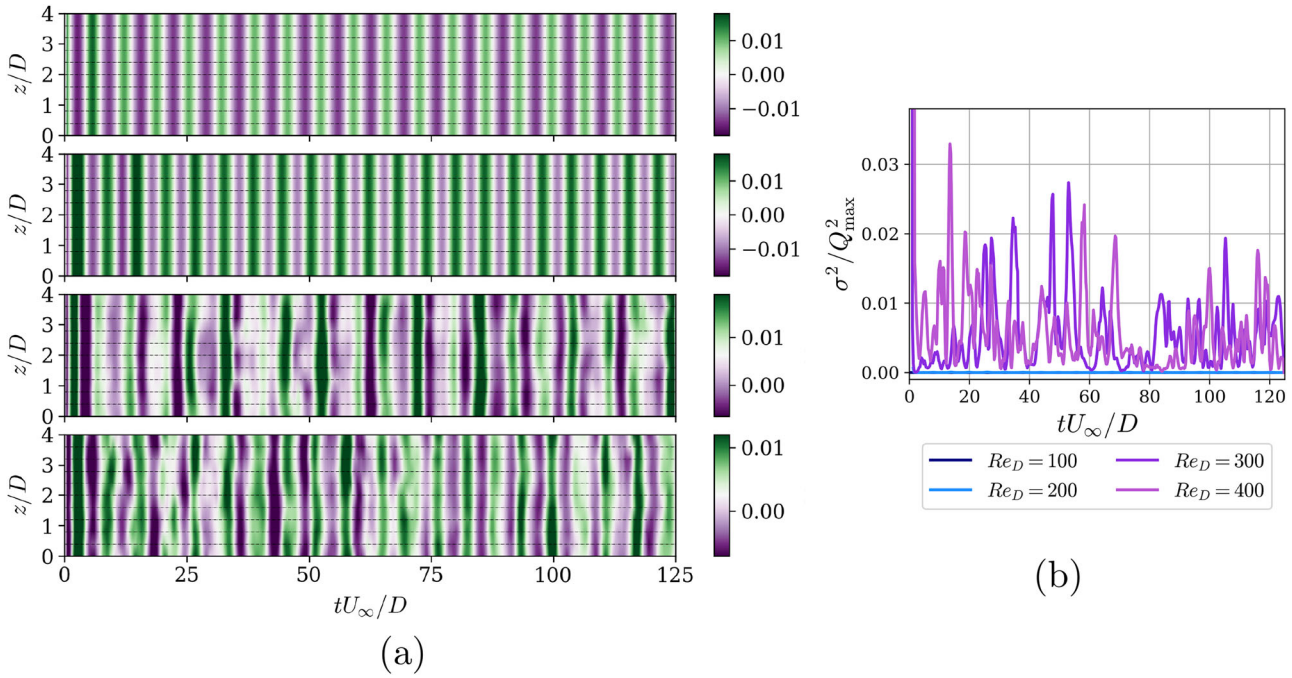

Fig. 3 | Evolution of the mass-flow rate associated with the jets in time and in the spanwise direction. a Mass-flow rate per unit width Q as a function of time for all jets individually, showing also their spanwise distribution for the DRL cases. From top to bottom: $Re_D = 100, 200, 300$, and 400. **b** Evolution in time of the variance of the mass-flow rate computed in z, $\sigma^2(t) = \frac{1}{n_{\text{jets}}}\sum_{i=1}^{n_{\text{jets}}}(Q_i(t) - \overline{Q}(t))^2$, for the different Reynolds numbers under study. Note that σ^2 is normalized by the squared peak Q values from each case, and $\sigma^2 = 0$ is obtained for $Re_D = 100$ and 200. The time is non-dimensionalized as tU_∞/D, where t is the physical time, U_∞ the freestream velocity, and D, the characteristic length, the cylinder diameter.

control begins to introduce spanwise variations. This is supported by the results in Fig. 3b, which shows the instantaneous variance of the mass flow rate in z over time for the various cases, denoted as σ^2. As mentioned in the Introduction, for $Re_D \geq 250$, the wake displays three-dimensional features, which are exploited by the DRL control to maximize drag-reduction rates.

During the exploration stage, for $Re_D = 100$ and 200, the agent did not find any spanwise-varying strategies that improved performance, suggesting that the wake is two-dimensional in these cases, favoring spanwise-uniform control strategies. In contrast, for $Re_D = 300$ and 400, flow patterns associated with transitional Re_D emerge, including spanwise structures of approximately one cylinder diameter ($\lambda_z/D = 1$), related to mode-B instabilities. Additionally, a basic SARL approach does not effectively exploit these local spanwise scales, while the MARL setup enables the use of these structures, as shown by the non-zero variance in the control in z.

Although it might seem plausible that a single-agent SARL approach could achieve spanwise-uniform strategies due to its simplicity, computational limitations make this approach unfeasible. Specifically, SARL faces the curse of dimensionality, requiring exponentially more trajectories to explore the state space thoroughly. This results in a significantly higher computational cost compared to the multi-agent DRL approach, which is more efficient in navigating complex control strategies.

In Fig. 4, we illustrate how the flow topology is influenced by the various drag-reduction strategies, on three representative phases: uncontrolled, transient, and stabilized control. The flow visualizations indicate that the control strategies based on DRL aim to enhance the spacing between successive vortical structures, resulting in a reduction of the vortex-shedding period T_k. Hence, mode-B instabilities are diminished when the control is applied, and the intensity of the vortex shedding is attenuated. These changes lead to a more organized wake structure, resembling the characteristic two-dimensional laminar wake. Figures 1b and 2a corroborate these findings, illustrating diminished oscillations during the controlled phase.

The observed reduction in unsteady structures within the recirculation bubble, or the increased absence of vortical activity in this region, can be associated with changes in the drag-reduction mechanisms. This interpretation is supported by statistical data presented later in this section in Fig. 5, which provides a more robust perspective on the flow dynamics.

Studying flow statistics offers deeper insight into the mechanisms employed by the DRL agent to discover flow-control strategies, particularly when analyzing the mean flow and the Reynolds stresses. To compute the latter, the Reynolds decomposition is used to decompose the flow variables (u) into time-averaged mean ($\overline{u}$) and fluctuating (u') components, $u = \overline{u} + u'$. In Fig. 5a, we observe the impact of the DRL-based control: the wake nearly doubles the recirculation-bubble length, delaying the wake-stagnation point by approximately one diameter in the streamwise direction. Instead of showing all cases, we only present $Re_D = 400$ as a representative case. Fig. 5b, c show how the wake also changes noticeably, being slightly wider but decaying much faster as we move downstream. Note that peaks in v fluctuations follow the same pattern, meaning that the counter-rotating vortices occur further as well. The pressure coefficient $C_p = 2(P - P_\infty)/(\rho U_\infty^2)$ around the cylinder is shown in Fig. 5d for the DRL-controlled case and the uncontrolled one. The back pressure increases by $\Delta\overline{C_{pb}} \approx 0.4$, which is directly related to the drag reduction mechanism.

The Reynolds stresses are presented in Fig. 5e, which shows that the peaks move downwards in the streamwise direction after applying the control, with only small changes in the vertical location. Additional analysis is provided in Fig. 5f, where the DRL-based control generally leads to the reduction of the peak magnitude in almost all the fluctuating quantities. In this case, all Re_D values are presented to elucidate that the same behavior occurs within this regime range. When considering fluctuations in the spanwise direction w', we notice a distinct pattern: an increase occurs at $Re_D = 300$, while a decrease is observed at higher Re_D values.

Additionally, we performed a proper-orthogonal-decomposition (POD) analysis[35] of the uncontrolled, PC and DRL-controlled cases at $Re_D = 400$ to better understand their flow physics. The PODs were performed with the pyLOM package[36] on the streamwise velocity and pressure

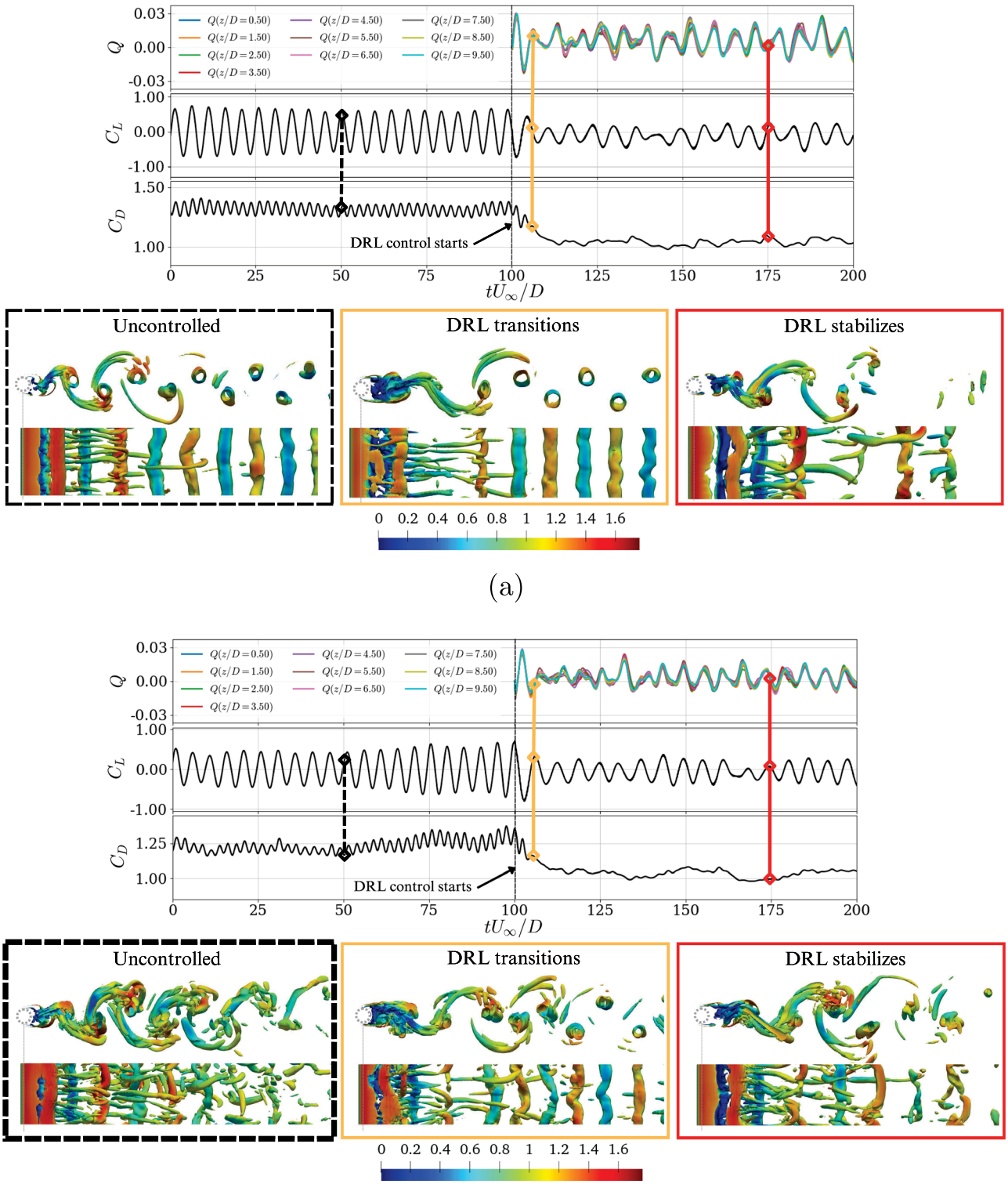

Fig. 4 | Visualizations of the flow coherent structures during DRL-based strategy exploitation. Temporal evolution of the flow at **a** $Re_D = 300$ and **b** $Re_D = 400$, from uncontrolled state to a stable DRL control. (Top) Non-dimensional mass-flow rate per unit width Q, lift coefficient C_L and drag coefficient C_D as a function of time; (bottom) snapshots showing vortical structures identified with the λ_2 criterion[55], where the isosurface $\lambda_2 D^2/U_\infty^2 = -0.5$ is shown for uncontrolled, transient, and DRL stabilized control states. The colors framing the flow visualizations correspond to the instants indicated in the temporal evolution of the relevant flow quantities.

fields over the last $40T_k$, with a sampling rate of about 25 snapshots per vortex-shedding.

Inspection of the frequency content of the temporal coefficients shown in Fig. 6a reveals that the DRL control acts selectively and less invasively. PC acts on a single frequency, resulting in a perturbation of the flow frequencies that can be seen in a more diffuse spectrum with a wider tail, thus indicating a stronger perturbation of the flow. In contrast, DRL control identifies and selects a wider range of frequencies to act on, resulting in a spectrum that is

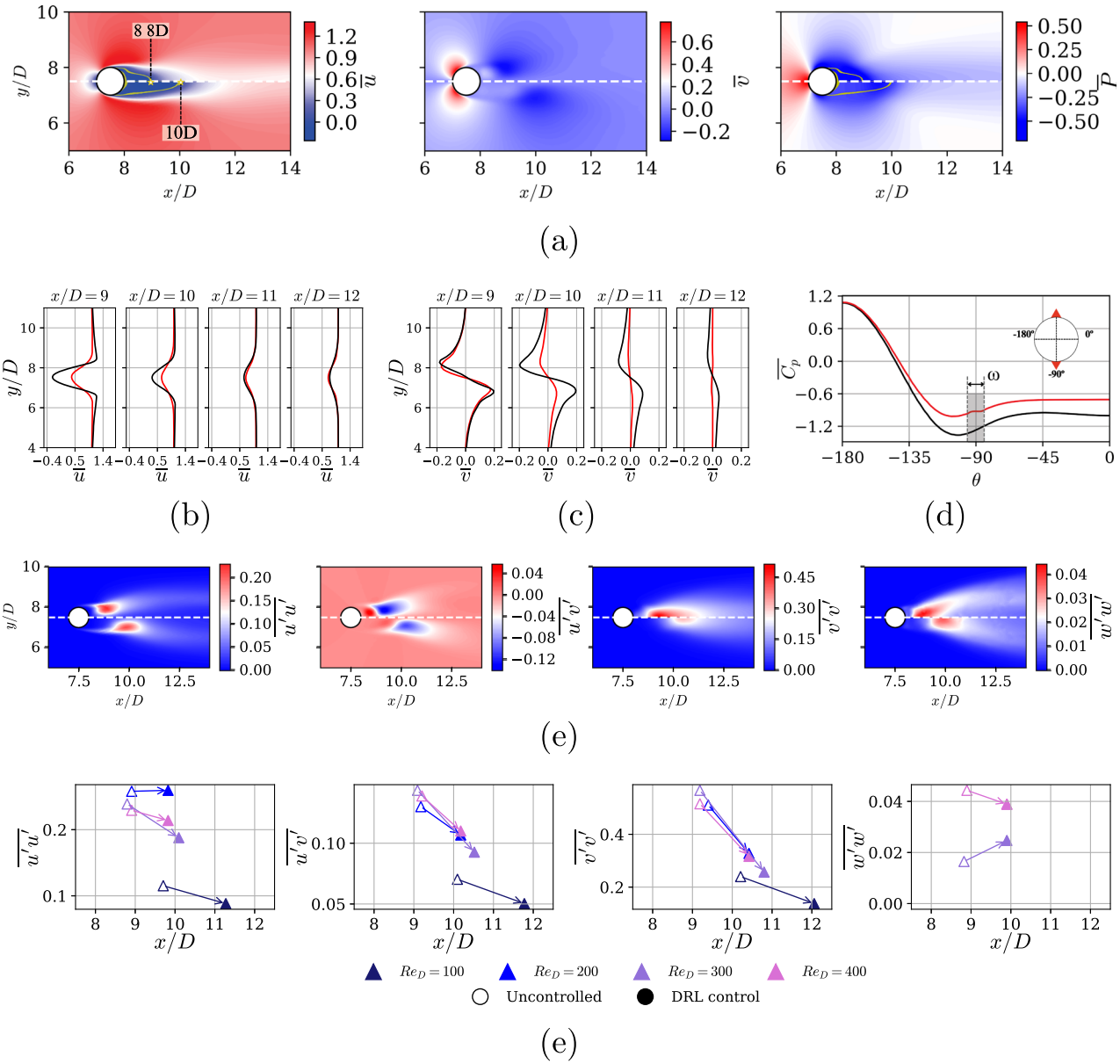

Fig. 5 | Mean flow and Reynolds stresses for the DRL-controlled cases. a Mean velocities and pressure fields for $Re_D = 400$, where (half top) is uncontrolled and (half bottom) is DRL-controlled flow. Yellow lines denote the regions where $\overline{u} = 0$ which indicate the wake-stagnation points also annotated with their streamwise location, x/D. Mean wake profiles of **b** $\overline{u}$ and **c** $\overline{v}$, as well as **d** mean pressure distribution $\overline{P}$ on the cylinder, respectively. We show (black) uncontrolled and (red) controlled cases. Note that the gray shaded area in **d** is the arc covered by the jet. **e** Reynolds stresses, from left to right, $\overline{u'u'}$, $\overline{u'v'}$, $\overline{v'v'}$ and $\overline{w'w'}$, where (half top) is uncontrolled and (half bottom) is DRL-controlled flow. **f** Peak values for Reynolds stresses and their x/D locations for all the Reynolds numbers under study. Note that $\overline{w'w'}$ for $Re_D = 100$ and 200 is not displayed because it is zero.

closer to that of the uncontrolled flow. As a result, the flow exhibits a narrower range of frequencies, yielding a less perturbed state.

Moreover, the streamwise component of the first POD mode is shown in Fig. 6b and the corresponding power-spectral densities are shown in Fig. 6c. This mode contains a significant contribution of the vortex-shedding for the uncontrolled flow case. In the PC there is, in addition to the identified vortex-shedding at $St|_{\mathrm{PC}} = 0.194$, as reported in Table 1, the contribution of the actuation frequency at $f_c|_{\mathrm{PC}} = 0.171$. For the DRL-controlled flow, we observe that the first mode closely resembles the control strategy found by the agent. Thus, DRL control is able to act on a wide range of frequencies, while PC only acts at a single frequency. This effectively alters the vortex-shedding frequency of the DRL-controlled flow to $St|_{\mathrm{DRL}} = 0.171$. In fact, DRL control actively works to modify the vortex-shedding frequency from the uncontrolled state, $St|_{\mathrm{Uncontrolled}} = 0.2$. As a consequence, there are significant differences in the flow characteristics. In PC, the resulting modes become highly disturbed on the far-wake, while the near-wake remains mostly unaltered, exhibiting a strong similarity compared with the uncontrolled case. Furthermore, for the PC, it is observed that the double peak of $f_c|_{\mathrm{PC}}$ and $St|_{\mathrm{PC}}$ coexist, indicating that the PC is not successfully modifying the shedding in the same way as the DRL-controlled system does.

On the other hand, in the DRL-controlled flow, the structures exhibit a more elongated wake structure and a larger recirculation bubble (measured from the end of the cylinder) of approximately $L_r/D = 2$, in contrast to the uncontrolled and PC recirculation bubbles of approximately $L_r/D = 0.8$ and $L_r/D = 0.9$, respectively. Subsequent modes exhibit more complex structures related to sub-harmonics and turbulent transition. The DRL-controlled flow is associated with modes that break the flow symmetries, even in the span-wise direction, suggesting a transition to three-dimensional flow at more energetic modes.

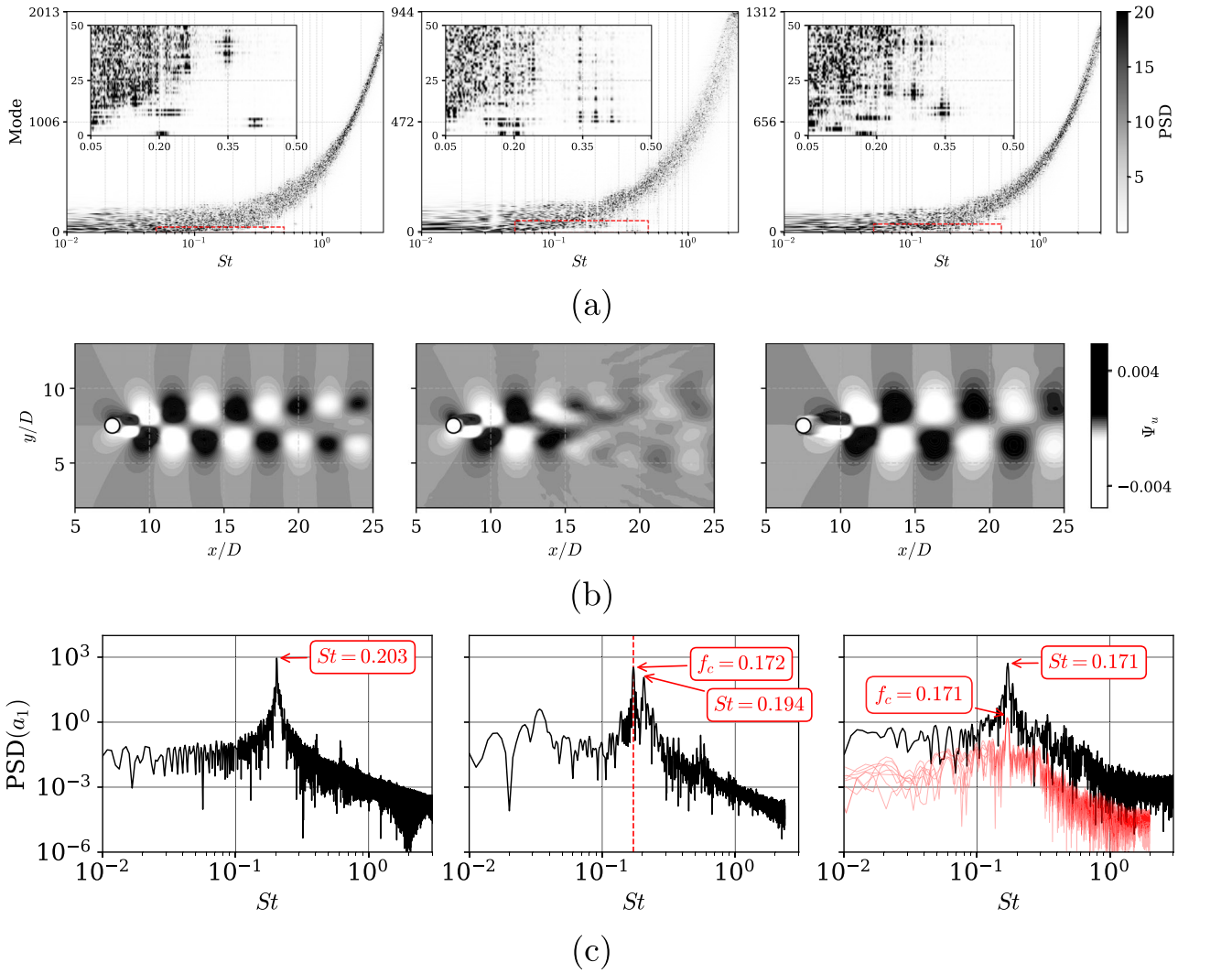

Fig. 6 | Proper-orthogonal-decomposition (POD) results for streamwise velocity u at $Re_D = 400$ comparing, from left to right: uncontrolled, PC, and DRL-controlled case. a Heatmap of the full spectrum for the POD modes and insert on top-left corner zooming into the first 50 modes within the dominant St. **b** Streamwise velocity for the first POD mode in the xy plane (homogeneous in the spanwise direction). **c** Power-spectral densities corresponding to the first POD mode, highlighting the shedding Strouhal number (St) and the characteristic frequency (f_c), which appears as a double peak in the PC case. Note that for PC, a dashed red vertical line $f_c|_{PC}$ is shown, while for DRL, the spectra of all Q_i are displayed in red with an offset to also highlight $f_c|_{DRL}$.

Discussion and conclusions

In this study, a MARL framework is coupled with a numerical solver to discover effective drag-reduction strategies by controlling multiple jets placed along the span of three-dimensional cylinders. We study cases at $Re_D = 100$, 200, 300, and 400, where wake transition from 2D to 3D is observed. All DRL-based control policies outperform the classical periodic control in this Re_D range. This is characterized by the emergence of spanwise instabilities, which the DRL agent can exploit to discover effective drag-reduction strategies. This is achieved by taking advantage of exploiting the underlying physics within pseudo-environments and optimizing the global problem involving multiple interactions in parallel. One of the main advantages of employing MARL is the capability to deploy trained agents across various cylinder lengths and numbers of actuators while ensuring consistency in the spanwise width of the jets (L_{jet}) and their corresponding pressure values as observation states (s_t). Note that the training focuses on symmetries and invariant structures. This would not be possible with SARL, which is restricted to a certain number of actuators (and also the corresponding algorithm limitations). MARL allows for cheaper training sessions in smaller and simplified computational domains, thereby speeding up the process, which is required to perform flow control in high-fidelity simulations.

These findings highlight the effectiveness of the DRL approach, which can discover flow-control strategies more sophisticated than those obtained with the classical periodic control, spanning wide ranges of frequencies and tackling different flow features in the wake. DRL-based control achieves a remarkable performance, reducing drag by 21% and 16.5% for $Re_D = 300$ and 400, respectively, outperforming PC strategies, which only achieve around 6% reduction for both Re_D. A POD of the DRL-controlled flow revealed the underlying physics of the actuation at $Re_D = 400$, resulting in a stable wake structure with a longer recirculation bubble.

In fact, addressing the curiosity of whether these DRL results could be simplified or reduced to a basic control-one resembling a periodic control, but already knowing the dominant frequencies from the DRL that yield such good performance and range of Q-was investigated. The idea was perhaps to find strategies that do not require solving such a high-dimensional problem. However, the results reported in the second part of the Supplementary Documentation Section 2 indicate that achieving drag reduction is not feasible without considering the full spectrum of both temporal and spatial frequencies utilized by MARL agents. The results show that alternative approaches do not even achieve a quarter of the performance on average. This wider spectrum of closed-loop strategies is essential for addressing the three-dimensionalities and

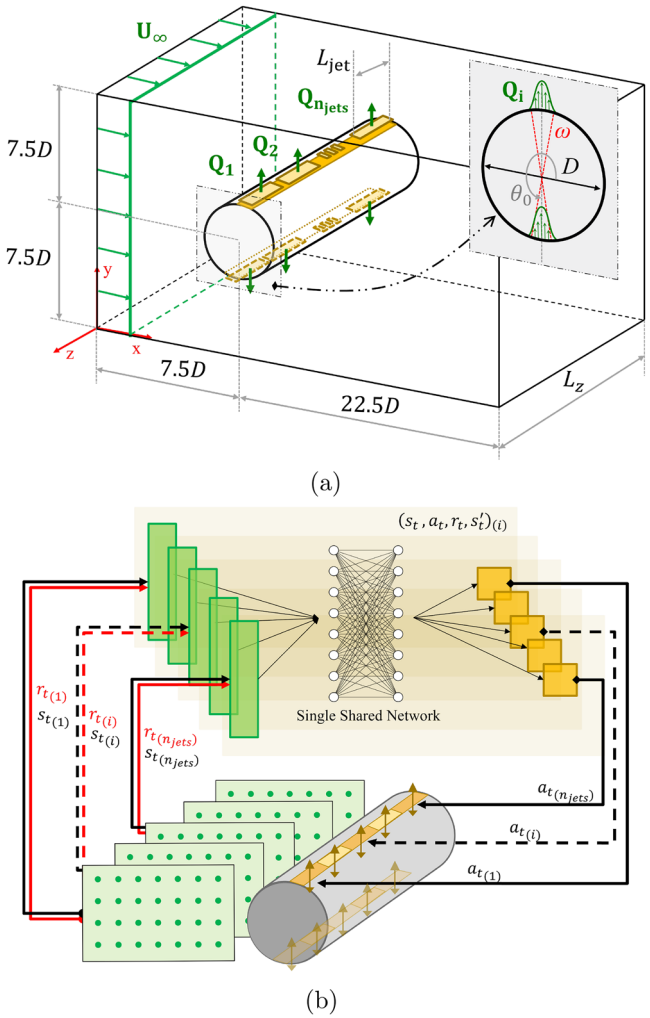

Fig. 7 | Conceptual visualization of the computational domain, the MARL framework, and a summary of parameters. a Schematic representation of the computational domain with cylinder diameter D as the reference length. Here ω is the jet width and θ_0 is the angular location of each jet. In green, we show the velocity condition for the inlet U_∞ and the sinusoidal profile in the jet azimuthal direction. **b** MARL framework applied to a three-dimensional cylinder equipped with distributed input distributed output (DIDO). Note that both **a** and **b** are not to scale.

transitional stages required to shift from an uncontrolled to a stabilized controlled state.

Furthermore, the results presented here represent the first training conducted in 3D cylinders using a MARL implementation. This study sets a new benchmark for the DRL community, potentially inspiring its application to more complex turbulent scenarios at higher Reynolds numbers and to other DIDO frameworks. Additionally, future research could improve training efficiency by implementing the approach in ref. 37, which explores the use of group invariants and positional encoding.

Methods

Problem configuration and numerical setup

The present study consists of a 3D cylinder exposed to a constant inflow in the streamwise direction. All domain lengths are non-dimensional, using the cylinder diameter D as the reference length. The geometry under consideration is presented in Fig. 7a. The computational domain has a streamwise length of $L_x/D = 30$, a height of $L_y/D = 15$, and a spanwise length of $L_z/D = 4$. The cylinder is centered at $(x/D, y/D) = (7.5, 7.5)$. Since the cylinder is considered to be infinitely long in the spanwise direction, we use periodic boundary conditions in z. Furthermore, we use a Dirichlet condition with a constant velocity U_∞ at the inlet. The top, bottom and outflow surfaces are outlets with imposed zero velocity gradient and constant pressure. The cylinder surfaces have no-slip and no-penetration conditions with zero velocity. The coordinate system origin is located at the front-face left-bottom corner. The last boundary conditions correspond to the cylinder actuators, which enable the control. The cylinder has a total of two sets of $n_{\text{jets}} = 10$ aligned synthetic jets that extend along its entire spanwise dimension, with a spanwise width $L_{\text{jet}}/D = 0.4$. This configuration enables having around two actuators per spanwise wavelength, which is known to be of the order of one diameter (mode-B). Future research could explore the optimal sizes and locations of these actuators for improved performance. Each set is placed at the top and bottom of the cylinder (at $\theta_0^{\text{top}} = 90^\circ$ and $\theta_0^{\text{bottom}} = 270^\circ$, respectively), defined as independent boundaries. The mass-flow rate can be changed by external actors (the DRL agent in this case, as discussed below). These actuators have an arc length in the xy plane of $\omega = 10°$, and no gap between the jets in z is considered. The jet-velocity profile is defined in terms of the angle θ and the desired mass-flow rate Q per unit width:

$$\| U_{\text{jet}}(Q,\theta) \| = Q\frac{\pi}{\rho D\omega}\cos\left(\frac{\pi}{\omega}(\theta-\theta_0)\right), \tag{2}$$

where $Q = \dot{m}/L_z$ and $|\theta-\theta_0| \in [-\omega/2, \omega/2]$, $\dot{m}$ is the mass flow rate. The absolute value of the jet velocity is projected into the x and y axes, since $\|U_{\text{jet}}\|$ corresponds to the radial cylinder direction. For each pseudo-environment, we set opposite action values within the pair of top and bottom jets, i.e., $Q_{90^\circ} = -Q_{270^\circ}$, in order to ensure the global zero-net mass flux. An earlier version of this setup was developed in ref. 38.

A conceptually similar control approach was reported in ref. 39, in particular, the one they denote as out phase. The out-phase approach consists of different constant mass-flow sinusoidal distributions along the spanwise direction with different wavelengths.

The transition from laminar to the emergence of the first three-dimensional instabilities in the cylinder wake occurs within a range of $Re_D = 200$ and 300[31,40–44]. The motivation of this research is to challenge the control system to discover the optimal strategies as the flow starts to become three-dimensional. By being able to manipulate scales with different spanwise wavelengths in the wake, the DRL-based control can learn effective mechanisms leading to a substantial reduction of the drag in the cylinder.

The numerical simulations are carried out by means of the numerical solver Alya, which is described in detail in ref. 45. The spatial discretization is based on the finite-element method and the incompressible Navier–Stokes equations:

$$\partial_t \boldsymbol{u} + (\boldsymbol{u}\cdot\nabla)\boldsymbol{u} - \nabla\cdot(2\nu\boldsymbol{\epsilon}) + \nabla p = \boldsymbol{f}, \tag{3}$$

$$\nabla\cdot\boldsymbol{u} = 0, \tag{4}$$

are integrated numerically, where $\boldsymbol{\epsilon}$ is a function of the velocity $\boldsymbol{u}$ which defines the velocity strain-rate tensor $\boldsymbol{\epsilon} = 1/2(\nabla\boldsymbol{u} + (\nabla\boldsymbol{u})^{\text{T}})$, $\boldsymbol{f}$ are the external body forces and ν is the kinematic viscosity. In Equation (4), the convective term $(\boldsymbol{u}\cdot\nabla)\boldsymbol{u}$ is expressed as a term conserving energy, momentum and angular momentum[46,47]. For the time discretization, a semi-implicit method is used where the convective term follows a second-order Runge–Kutta scheme and a Crank–Nicholson scheme is used for the diffusive term[48]. To select the appropriate time step, Alya uses an eigenvalue-based time-integration scheme[49]. Then, for each time step, the numerical solution of these equations is computed. Drag and lift forces (F_x and F_y, respectively) are obtained by integration over the cylinder surface $\boldsymbol{s}$:

$$\boldsymbol{F} = \int(\boldsymbol{\varsigma}\cdot\boldsymbol{n})\cdot\boldsymbol{e}_j \mathrm{d}\boldsymbol{s}, \tag{5}$$

where $\boldsymbol{\varsigma}$ is the Cauchy stress tensor, $\boldsymbol{n}$ the unit vector normal to the surface and $\boldsymbol{e}_j$ is a unit vector with the direction of the main flow velocity for F_x and the perpendicular cross-flow direction to it for F_y.

Multi-agent reinforcement learning

In the present work, we implemented a DRL framework using Tensorforce libraries[50]. DRL is very well suited for unsteady flow-control problems as it provides the possibility to dynamically interact with an environment, being able to dynamically set the actuation based on the varying flow state. We use the proximal policy-optimization (PPO) algorithm[51], which is a policy-gradient approach based on a surrogate loss function for policy updates to prevent drastic drops in performance. This algorithm demonstrates robustness, as it is relatively tolerant to initial hyperparameter settings and performs well across a wide variety of RL tasks without requiring extensive tuning. In this context, ref. 52 is referenced, where the effects of parallel environments, action frequency update T_a/T_k, and the smoothing law are studied. Also, ref. 53, where in Appendix E, different s_t had consistent results.

The neural-network architecture consists of two dense layers of 512 neurons each. The batch size, i.e., the total number of experiences that the PPO agent uses for each gradient-descent iteration, is set to 80, which is larger than the typical values used in 2D trainings[53,54], but sufficiently small to efficiently update the neural-network parameters. If there is a discrepancy between the batch size and the environments running at the same time, there is a risk of wasting information that will not be captured by the agent. The limitation is that we have 10 actuators per environment, $n_{\text{jets}} = 10$, and we need 10 streamed experiences, which will be synchronized, so we have to work with a total of $n_{\text{jets}} \times n_{\text{envs}}$ set of experiences. A streamed experience consists of a set of states, actions, rewards, and the predicted state that the agent expects to achieve. It is denoted as $(s_t, a_t, r_t, s')_i$ for each pseudo-environment, and each of the Reynolds numbers under consideration has its own agent and policy.

Previous work on 2D cylinders employed SARL to implement various training stages. However, if the action space needs to handle multiple jets simultaneously—such as in the present 3D cylinder setup with distributed input forcing and a distributed output reward (the so-called DIDO scheme) where SARL is not viable. Unlike SARL, the MARL framework mitigates the curse of dimensionality by exploiting symmetries and training agents in local pseudo-environments. This approach makes high-dimensional control tasks more tractable by breaking them into smaller domains, allowing agents to maximize local rewards effectively. Furthermore, by distributing both the input and output spaces across multiple agents, the exploration space becomes more manageable. This not only accelerates convergence during training but also ensures that the training process remains computationally feasible for such high-dimensional challenges.

It is important to note that earlier attempts using SARL, although not detailed in this study, were unsuccessful in achieving the desired performance, similar to the findings in Rayleigh–Bénard convection reported in ref. 25. In both cases, SARL struggled with handling global information effectively, leading to suboptimal control strategies. The limitations of the single-agent approach became evident even in simpler scenarios, where it failed to generalize across complex, high-dimensional environments. In contrast, MARL not only overcomes these challenges but also demonstrates superior performance by leveraging localized coordination among agents. This ability to process distributed input and optimize distributed rewards makes MARL a much more effective framework, even in cases where SARL had previously been considered feasible.

Figure 7b can help to understand the forthcoming explanation of the MARL setup. All the agents share the same neural-network weights, which is a key factor in accelerating the training process. Note that each agent is coupled to a pair of jets that actuate independently from the others through the training process.

The observation state s_t provided to the agent consists of partial pressure values along the domain. This information is composed of three slices, each containing 99 pressure values, which are aligned with the corresponding jet in z/D coordinates and separated by $L_z/30D$. The probes or pressure values are concentrated in the wake and near-cylinder regions, allowing the agent to effectively exploit the spanwise pressure gradients.

In prior work in ref. 38, various configurations were tested, including changes to the spanwise location and number of slices, to evaluate their impact on the performance of the drag reduction algorithm. These tests were crucial. The configuration chosen for this study was selected because it consistently provided the best overall performance across the evaluated scenarios.

The total reward $R(t, i_{\text{jet}})$ defined in Equation (6) is expressed as a sum of the local, r_{local}, and global, r_{global}, rewards that correspond to each jet i_{jet}. The heuristic scalar K_R adjusts the values within the range $[-1, 1]$, and β balances the local and global rewards; note that a value of $\beta = 0.8$ is used in this work. This means that 80% of the weight is assigned to the local value, while the remaining 20% accounts for the global value. Based on our experience, the parameter β also acts as a smoother for the reward signal. If β is too high, the signal can become noisy, whereas a lower value significantly reduces fluctuations. Thus, in terms of learning a policy, the control authority, understood as the ability to influence the behavior of the system given feedback, can also be significantly influenced.

We acknowledge that we did not experimentally test a wide range of β values, as this would have been computationally prohibitive. However, this limitation motivates future research to better understand the sensitivity of β and its role in balancing local and non-local information. Importantly, β plays a key role in facilitating coordination between agents, enabling neighboring agents to exchange meaningful information. This coordination is critical because what is beneficial for local performance may sometimes conflict with global objectives, and vice versa. By defining $\beta = 0.8$, the framework seeks to find a balance that harmonizes local and global priorities.

The rewards r_t, defined in Equation (7), are functions of the aerodynamic force coefficients C_D and C_L (note that $\overline{C_{D_b}}$ is the uncontrolled averaged drag in a stationary state). The user-defined parameter α is a lift penalty, and in this study, we considered $\alpha = 0.6$, which provides a good trade-off between ensuring symmetric strategies without excessively restricting the exploration process. The latter is essential to avoid undesired asymmetric strategies that favor a reduction of the component parallel to the incident velocity (drag) towards the perpendicular one (positive or negative lift). This phenomenon is commonly referred to as the axis-switching phenomenon.

$$R(t, i_{\text{jet}}) = K_R\left[\beta r_{\text{local}}(t, i_{\text{jet}}) + (1-\beta) r_{\text{global}}(t)\right], \tag{6}$$

$$r(t, i_{\text{jet}}) = \overline{C_{D_b}} - C_D(t, i_{\text{jet}}) - \alpha|C_L(t, i_{\text{jet}})|, \tag{7}$$

$$\textit{where} \quad C_D = \frac{2F_x}{\rho A_f U_\infty^2} \quad \text{and} \quad C_L = \frac{2F_y}{\rho A_f U_\infty^2}. \tag{8}$$

The aerodynamic forces involve the frontal area $Af = DL_z$ from the local pseudo-environment surfaces for $C_{D_{\text{local}}}$ and the whole cylinder for $C_{D_{\text{global}}}$.

The interactions between the agent and the physical environment are denoted as actions a_t, and they influence the system during T_a time units. We update the jet boundary conditions using Equation (2). The shift in time between actions, $Q_t \rightarrow Q_{t+1}$ is managed through an exponential function. The smooth transition diminishes the appearance of sudden discontinuities, which can spoil a training process.

The DRL library outputs values in the range $a_t \in [-1, 1]$, requiring rescaling as $Q = a_t Q_{\max}$ to introduce the magnitude of actuation later in Equation (2). To balance exploration, learning efficiency, and numerical stability, $Q_{\max} = 0.176$ was selected based on our experience with DRL for flow control. This value ensures meaningful exploration without introducing excessively large actuations that could destabilize the learning process or the CFD solver. If $Q_{\max}$ were set too small, the exploration space would be overly constrained, potentially leading to premature policy saturation and suboptimal solutions. On the other hand, excessively large $Q_{\max}$ values could result in an expanded exploration space that delays convergence or imposes boundary conditions that challenge the CFD solver's stability. Notably, $Q_{\max} = 0.176$ corresponds to twice the values used in the 2D cylinder setups[54], reflecting adjustments for the current configuration and objectives.

Table 2 | Summary of the main parameters for the present DRL framework

Parameter	Value
n_{jets}	10
s_t size	297 (99 in 3 *xy*-slices)
s_t variable	Pressure
Q_{max}	0.176
Reward scalar K_R	5
Lift penalty α	0.6
Reward local weight β	0.8
Action duration $T_a[tU_\infty/D]$	0.25
Actions per episode	120
Time-smoothing function	Exponential
Batch size M	60
Epoch for optimizer	25
Strategy optimize policy	PPO
Architecture (networks)	2
Network 1 & 2 (type)	Fully connected layer
Network 1 & 2 (size)	512 neurons
Learning rate α_θ and α_ϕ	0.001
Likelihood ratio clipping ϵ	0.2
Discount factor γ	0.99
Optimizer type	Adam
Entropy regularization λ	0.01

Certain parameters in the DRL configuration are closely tied to the fluid mechanics problem under consideration. The episode duration is specifically defined to include at least six vortex-shedding periods ($T_k = 1/St$). We set $T_a < 0.05T_k$, based on the experience gathered with previous studies[12,53]. This allows sufficient time between actions to produce an effect on the flow. Note that if the time between actions is too short, there will be noise in the training process, and it will become difficult to converge. On the other hand, if this is too large, the agent will not be able to control the smaller-scale structures associated with shorter time scales. Thus, a total of 120 actuations per episode is deemed sufficient for evaluating the cumulative reward. It is noteworthy that each episode starts from an uncontrolled converged state of the problem. This corresponds to what happens during training, but when we evaluate the DRL model in exploitation mode (also denoted as a deterministic mode), we make the episodes 4 times longer to ensure statistical convergence.

Table 2 collects and summarizes all the main parameters required to set up this DRL framework, which is coupled with a CFD solver, many of which were discussed in this section. Note that the agent hyperparameters like ϵ or α_θ, are also included but are not discussed in detail in this study.

Note that we also compare the DRL-based control with results from the classical periodic control. The latter is chosen with the same jet flow rate as that of the DRL, and the frequency is chosen based on a parametric analysis of the frequency around the vortex-shedding frequency of the wake. We selected the frequency yielding the highest drag reduction.

Extensive work documented in ref. 38 was carried out to adjust the MARL framework and the communications setup. For instance, the definition of s_t is the result of a compromise between computational practicality and physical relevance. The spanwise wavelength of the structures in the wake also helped to define the spacing of the *xy* planes defining the system state. Note that the number of data points used for this state correlates with the number of weights calculated for the first fully connected layer of the neural network.

Data availability

The CFD case data used in this study are available from the corresponding author upon reasonable request, due to their large size and the need for specific postprocessing tools associated with the Alya multi-physics solver. However, the trained Tensorforce models for the agents are included in the aforementioned code repository and are openly accessible.

Code availability

The code used in this study is available at https://github.com/KTH-FlowAI/Flow-control-3Dcylinders-via-MARL. The repository contains all necessary scripts to reproduce the results presented in the manuscript. The code is provided under the *MIT License*.

Received: 11 May 2024; Accepted: 3 June 2025;
Published online: 18 June 2025

References

1. Choi, H., Jeon, W.-P. & Kim, J. Control of flow over a bluff body. *Annu. Rev. Fluid Mech.* **40**, 113–139 (2008).
2. Raymer, D. P. *Aircraft design: a conceptual approach*. AIAA education series, 4th edn. (American Institute of Aeronautics and Astronautics, 2006).
3. Whitcomb, R. T. A design approach and selected wind tunnel results at high subsonic speeds for wing-tip mounted winglets. *Technical Report*, https://ntrs.nasa.gov/api/citations/19760019075/downloads/19760019075.pdf (1976).
4. Lin, J. C. Review of research on low-profile vortex generators to control boundary-layer separation. *Prog. Aerosp. Sci.* **38**, 389–420 (2002).
5. Siddiqui, N., Asrar, W. & Sulaeman, E. Literature review: biomimetic and conventional aircraft wing tips. *Int. J. Aviat., Aeronautics, Aerosp.* **4**, 5–8 (2017).
6. Guerrero, J. E., Maestro, D. & Bottaro, A. Biomimetic spiroid winglets for lift and drag control. *C. R. Mécanique* **340**, 67–80 (2012).
7. Vinuesa, R., Brunton, S. L. & McKeon, B. J. The transformative potential of machine learning for experiments in fluid mechanics. *Nat. Rev. Phys.* **5**, 536–545 (2023).
8. Vinuesa, R. & Brunton, S. L. Enhancing computational fluid dynamics with machine learning. *Nat. Comput. Sci.* **2**, 358–366 (2022).
9. Le Clainche, S. et al. Improving aircraft performance using machine learning: a review. *Aerosp. Sci. Technol.* **138**, 108354 (2023).
10. Garnier, P. et al. A review on deep reinforcement learning for fluid mechanics. *Comput. Fluids* **225**, 104973 (2021).
11. Silver, D. et al. Mastering the game of go with deep neural networks and tree search. *Nature* **529**, 484–503 (2016).
12. Tang, H., Rabault, J., Kuhnle, A., Wang, Y. & Wang, T. Robust active flow control over a range of Reynolds numbers using an artificial neural network trained through deep reinforcement learning. *Phys. Fluids* **32**, 053605 (2020).
13. Xu, H., Zhang, W., Deng, J. & Rabault, J. Active flow control with rotating cylinders by an artificial neural network trained by deep reinforcement learning. *J. Hydrodynamics* **32**, 254–258 (2020).
14. Paris, R., Beneddine, S. & Dandois, J. Robust flow control and optimal sensor placement using deep reinforcement learning. *J. Fluid Mech.* **913**, A25 (2021).
15. Li, J. & Zhang, M. Reinforcement-learning-based control of confined cylinder wakes with stability analyses. *J. Fluid Mech.* **932**, A44 (2022).
16. Ren, F., Rabault, J. & Tang, H. Applying deep reinforcement learning to active flow control in weakly turbulent conditions. *Phys. Fluids* **33**, 037121 (2021).
17. Fan, D., Yang, L., Wang, Z., Triantafyllou, M. S. & Karniadakis, G. E. Reinforcement learning for bluff body active flow control in experiments and simulations. *Proc. Natl Acad. Sci.* **117**, 26091–26098 (2020).

18. Chatzimanolakis, M., Weber, P. & Koumoutsakos, P. Learning in two dimensions and controlling in three: Generalizable drag reduction strategies for flows past circular cylinders through deep reinforcement learning. *Phys. Rev. Fluids* **9**, 043902 (2024).
19. Vinuesa, R., Lehmkuhl, O., Lozano-Durán, A. & Rabault, J. Flow control in wings and discovery of novel approaches via deep reinforcement learning. *Fluids* **7**, 62 (2022).
20. Chen, W., Wang, Q., Yan, L., Hu, G. & Noack, B. R. Deep reinforcement learning-based active flow control of vortex-induced vibration of a square cylinder. *Phys. Fluids* **35**, 053610 (2023).
21. Guastoni, L., Rabault, J., Schlatter, P., Azizpour, H. & Vinuesa, R. Deep reinforcement learning for turbulent drag reduction in channel flows. *Eur. Phys. J. E* **46**, 27 (2023).
22. Yan, X., Zhu, J., Kuang, M. & Wang, X. Aerodynamic shape optimization using a novel optimizer based on machine learning techniques. *Aerosp. Sci. Technol.* **86**, 826–835 (2019).
23. Viquerat, J. et al. Direct shape optimization through deep reinforcement learning. *J. Comput. Phys.* **428**, 110080 (2021).
24. Keramati, H., Hamdullahpur, F. & Barzegari, M. Deep reinforcement learning for heat exchanger shape optimization. *Int. J. Heat. Mass Transf.* **194**, 123112 (2022).
25. Vignon, C. et al. Effective control of two-dimensional Rayleigh–Bénard convection: Invariant multi-agent reinforcement learning is all you need. *Phys. Fluids* **35**, 065146 (2023).
26. Kurz, M., Offenhäuser, P. & Beck, A. Deep reinforcement learning for turbulence modeling in large eddy simulations. *Int. J. Heat. Fluid Flow.* **99**, 109094 (2023).
27. Novati, G., de Laroussilhe, H. L. & Koumoutsakos, P. Automating turbulence modeling by multi-agent reinforcement learning. *Nat. Mach. Intell.* **3**, 87–96 (2021).
28. Beck, A. & Kurz, M. Toward discretization-consistent closure schemes for large eddy simulation using reinforcement learning. *Phys. Fluids* **35**, 125122 (2023).
29. Bae, H. J. & Koumoutsakos, P. Scientific multi-agent reinforcement learning for wall-models of turbulent flows. *Nat. Commun.* **13**, 1443 (2022).
30. Wang, Z., Fan, D., Jiang, X., Triantafyllou, M. S. & Karniadakis, G. E. Deep reinforcement transfer learning of active control for bluff body flows at high reynolds number. *J. Fluid Mech.* **973**, A32 (2023).
31. Williamson, C. Vortex dynamics in the cylinder wake. *Annu. Rev. Fluid Mech.* **28**, 477–539 (1996).
32. Barkley, D. & Henderson, R. D. Three-dimensional Floquet stability analysis of the wake of a circular cylinder. *J. Fluid Mech.* **322**, 215–241 (1996).
33. Brunton, S. L. & Noack, B. R. Closed-loop turbulence control: progress and challenges. *Appl. Mech. Rev.* **67**, 050801 (2015).
34. Vignon, C., Rabault, J. & Vinuesa, R. Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions. *Phys. Fluids* **35**, 031301 (2023).
35. Lumley, J. L. The structure of inhomogeneous turbulent flows. In *Yaglom, A.M. and Tartarsky, V.I., Eds., Atmospheric Turbulence and Radio Wave Propagation* 166–177 (1967).
36. Eiximeno, B. et al. Pylom: a HPC open source reduced order model suite for fluid dynamics applications. *Comput. Phys. Commun.* **308**, 109459 (2025).
37. Jeon, J. et al. Advanced deep-reinforcement-learning methods for flow control: group-invariant and positional-encoding networks improve learning speed and quality. *arxiv* https://arxiv.org/abs/2407.17822 (2024).
38. Suárez, P. et al. Active flow control for three-dimensional cylinders through deep reinforcement learning. In: *14th International ERCOFTAC Symposium on Engineering, Turbulence, Modelling and Measurements: 6th-8th September 2023, Barcelona, Spain: proceedings*. (2023).
39. Kim, J. & Choi, H. Distributed forcing of flow over a circular cylinder. *Phys. Fluids* **17**, 033103 (2005).
40. Bays-Muchmore, B. & Ahmed, A. On streamwise vortices in turbulent wakes of cylinders. *Phys. Fluids A: Fluid Dyn.* **5**, 387–392 (1993).
41. Bloor, M. S. The transition to turbulence in the wake of a circular cylinder. *J. Fluid Mech.* **19**, 290–304 (1964).
42. Karniadakis, G. E. & Triantafyllou, G. S. Frequency selection and asymptotic states in laminar wakes. *J. Fluid Mech.* **199**, 441–469 (1989).
43. Karniadakis, G. E. & Triantafyllou, G. S. Three-dimensional dynamics and transition to turbulence in the wake of bluff objects. *J. Fluid Mech.* **238**, 1–30 (1992).
44. Norberg, C. An experimental investigation of the flow around a circular cylinder: influence of aspect ratio. *J. Fluid Mech.* **258**, 287–316 (1994).
45. Vázquez, M. et al. Alya: multiphysics engineering simulation toward exascale. *J. Comput. Sci.* **14**, 15–27 (2016).
46. Charnyi, S., Heister, T., Olshanskii, M. A. & Rebholz, L. G. On conservation laws of navier–stokes galerkin discretizations. *J. Comput. Phys.* **337**, 289–308 (2017).
47. Charnyi, S., Heister, T., Olshanskii, M. A. & Rebholz, L. G. Efficient discretizations for the EMAC formulation of the incompressible Navier–Stokes equations. *Appl. Numer. Math.* **141**, 220–233 (2019).
48. Crank, J. & Nicolson, P. A practical method for numerical evaluation of solutions of partial differential equations of the heat-conduction type. *Math. Proc. Camb. Philos. Soc.* **43**, 50–67 (1947).
49. Trias, F. & Lehmkuhl, O. A self-adaptive strategy for the time integration of Navier-Stokes equations. *Numer. Heat. Transf. B. Fundam.* **60**, 116–134 (2011).
50. Schaarschmidt, M., Kuhnle, A. & Fricke, K. TensorForce: a TensorFlow library for applied reinforcement learning. Available at: https://github.com/reinforceio/tensorforce (2017).
51. Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. *Preprint at:* https://arxiv.org/abs/1707.06347 (2017).
52. Rabault, J. & Kuhnle, A. Accelerating deep reinforcement learning strategies of flow control through a multi-environment approach. *Phys. Fluids* **31**, 094105 (2019).
53. Rabault, J., Kuchta, M., Jensen, A., Reglade, U. & Cerardi, N. Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *J. Fluid Mech.* **865**, 281–302 (2019).
54. Varela, P. et al. Deep reinforcement learning for flow control exploits different physics for increasing Reynolds number regimes. *Actuators* **11**, 359 (2022).
55. Jeong, J. & Hussain, F. On the identification of a vortex. *J. Fluid Mech.* **285**, 69–94 (1995).

Acknowledgements

This study was enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at PDC, KTH Royal Institute of Technology. R.V. acknowledges financial support from the ERC grant no. 2021-CoG-101043998, DEEPCONTROL. Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Author contributions

S.P.: methodology, software, validation, investigation, writing–original draft and visualization. A-Á.F., R.J., M.A. & F.B.: Methodology, software, and writing–review & editing. L.O.: funding acquisition, supervision, and writing–review & editing. V.R.: conceptualization, project definition, methodology, resources, writing–original draft, supervision, project administration and funding acquisition.

Funding
Open access funding provided by Royal Institute of Technology.

Competing interests
The authors declare no competing interests.

Inclusion and Ethics
The authors affirm that the research presented in this manuscript respects principles of inclusion, diversity, and ethics. Collaborative contributions from all authors were valued equally, and the research was conducted following ethical standards relevant to global scientific research. No biases related to race, gender, geographic location, or socioeconomic status influenced the study.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s44172-025-00446-x.

Correspondence and requests for materials should be addressed to Pol Suárez or Ricardo Vinuesa.

Peer review information *Communications Engineering* thanks Feng Ren, Alistair Revell, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: [Miranda Vinay, Anastasiia Vasylchenkova]. Peer reviewer reports are available.