



MOX-Report No. 40/2023

Reducing the Drag of a Bluff Body by Deep Reinforcement Learning

Ballini, E.; Chiappa, A.S.; Micheletti, S.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<https://mox.polimi.it>

Reducing the Drag of a Bluff Body by Deep Reinforcement Learning

E. Ballini[#], A. S. Chiappa[†], S. Micheletti[#]

May 5, 2023

[#] MOX – Modellistica e Calcolo Scientifico, Dipartimento di Matematica
Politecnico di Milano, Piazza L. da Vinci 32, I-20133 Milano, Italy
{enrico.ballini, stefano.micheletti}@polimi.it

[†] EPFL, Brain Mind Institute, Route Cantonale, CH-1015, Lausanne, Switzerland
alberto.chiappa@epfl.ch

Abstract

We present a deep reinforcement learning approach to a classical problem in fluid dynamics, i.e., the reduction of the drag of a bluff body. We cast the problem as a discrete-time control with continuous action space: at each time step, an autonomous agent can set the flow rate of two jets of fluid, positioned at the back of the body. The agent, trained with Proximal Policy Optimization, learns an effective strategy to make the jets interact with the vortexes of the wake, thus reducing the drag. To tackle the computational complexity of the fluid dynamics simulations, which would make the training procedure prohibitively expensive, we train the agent on a coarse discretization of the domain. We provide numerical evidence that a policy trained in this approximate environment still retains good performance when carried over to a denser mesh. Our simulations show a considerable drag reduction with a consequent saving of total power, defined as the sum of the power spent by the control system and of the power of the drag force, amounting to 40% when compared to simulations with the reference bluff body without any jet. Finally, we qualitatively investigate the control policy learnt by the neural network. We can observe that it achieves the drag reduction by learning the frequency of formation of the vortexes and activating the jets accordingly, thus blowing them away off the rear body surface.

1 Introduction

The reduction of the drag force is a problem of paramount importance in fluid dynamics, because of its ubiquity in aeronautical, naval, and land transport applications. The drag force, acting in the opposite direction to the body’s motion, is often the principal source of power consumption. The problem has been deeply studied both for aerodynamic bodies, such as aircraft surfaces, and bluff bodies, i.e., those bodies with a compact geometry that show conspicuous recirculation areas. In this work, we propose a method for drag reduction with deep reinforcement learning. We consider a 2D bluff body at high Reynolds number, simulating the flow field with unsteady Computational Fluid Dynamics (CFD) (Fig. 1). The geometry of the body is rectangular, with rounded front edges and with the addition of two small curved edges at the back that alter the initial geometry in a negligible way (Fig. 2). The drag of a bluff body is mainly caused by the vast wake behind it because it strongly influences the pressure distribution around the body. Therefore, our goal is to reduce the drag force by interacting with the wake. This is done by introducing two small nozzles, positioned on the rear surface, that can interact with the wake by emitting fluid at a desired flow rate. The two jets of fluid, coming out of the nozzles, blow on the two small curved edges positioned at the rear corners. We cast the control of the fluid injection rate as a continuous control problem, which can be solved through reinforcement learning. An autonomous agent must learn a policy to optimally control the two nozzles. We remark that the learning is purely simulation-driven, as it does not require any a-priori knowledge of the equations governing the system. The agent learns through experience the function linking its current observation of the environment to the action that maximally reduces the drag force. The observation consists in the real-time measurement of the pressure on the base surface of the bluff body, at 12 fixed points. In spite of the limited amount of information made available to the agent, the trained control policy successfully reduces the drag force on the bluff body. Compared to a simulation without active jet control, the net drag power saving amounts to 40%.

An analysis of the control policy learned by the neural network reveals how the jets interact with the wake vortexes to highly reduce the drag force. We found that the agent ignites the jet on the opposite side where the vortex is being generated, pushing its formation downstream. This mechanism increases the base pressure, which is the main cause of the drag reduction.

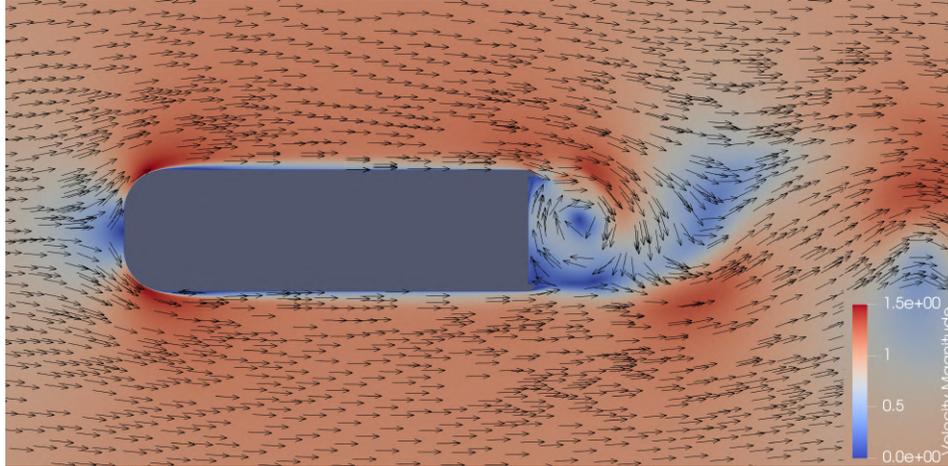


Figure 1: Snapshot of the flow around the body. The arrows represent the velocity field and the color map the associated magnitude.

2 Related works

A long list of approaches to reduce the drag of bluff bodies have been proposed. We can find an overview in [9, 14]. Examples of some particular applications are: surface modifications to control the separation point [21, 29, 52], porous surfaces [19, 26], addition of other bodies or appendages that interact with the main body [17, 11, 51, 24, 41, 23, 39], plasma actuators [8, 25], and closed-loop control of back oscillating flaps [5]. Many studies concern the use of suitably positioned jets of fluid. The flow can be ejected at constant velocity [60], at variable velocity with non-zero mean flow rate [20, 58, 54, 36], or at variable velocity with null mean flow rate [43]. The literature regarding controlled jets is less abundant. Some important studies concern the control of the wake through variable jets [22, 30, 12]. A relevant practical application of the drag reduction of a bluff body is found in heavy vehicles because lower drag means lower fuel consumption. Indeed, solutions focused on both the front and rear have been found for these vehicles [48, 10]. Also Machine Learning (ML) has found applications in different branches of fluid dynamics. For an overview, we refer to [47, 6, 27]. In particular, Reinforcement Learning (RL) [53] has also been applied in the fluid dynamics field [18, 45]. Some particular applications are: PDE control [7, 16], flow control [46, 1, 15], control of the movement of objects immersed in a fluid, such as fish-like swimmers [35], microswimmers [34], Zermelo problem [2], drones e gliders [4, 44], as well as more general applications as shape optimization [57], and the creation of numeri-

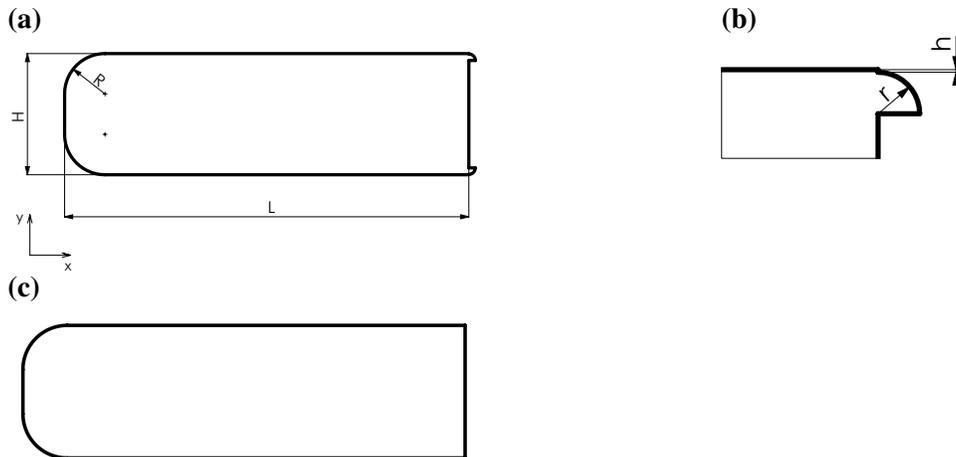


Figure 2: Schematic representation of the bluff bodies used for the fluid dynamics simulations. **(a)** Geometry of the bluff body with controlled jets. **(b)** Curved edge and slit detail. **(c)** Geometry of the bluff body clean used as reference for the drag reduction evaluation.

cal methods to approximate the solution to differential equations [59]. To the best of the authors' knowledge, the application of RL presented in [61] is the closest to ours. Indeed, the authors of [61], by providing the RL agent with an accurate measurement of the fluid flow, managed to reduce the oscillation of a cylinder immersed in a low-Reynolds number flow.

3 Fluid dynamics model

The numerical experiments simulate the flow around a 2D bluff body of height $H = 1$ and length $L = 3.3$, as shown in Fig. 2. All quantities are dimensionless. The front is rounded with a constant radius of $R = 0.3$ in order to avoid large separations after the leading edge. The back jets are blown by two slits of height $h = 0.03$. The jets blow on two 90° circular edges of radius $r = 0.05$. The Reynolds number, based on the height of the bluff body, is $Re = \frac{HU_\infty}{\nu} = 20000$, where U_∞ is the velocity at infinity, and ν is the kinematic viscosity. We evaluate the drag reduction resulting from the simulations against a reference clean bluff body depicted in Fig. 2. It is nearly the same as the presented bluff body, with the only difference that it has neither jets nor curved edge on the back, so the back is totally plain. The fluid motion is modeled by the URANS equations [42, 3] and the $k - \omega$ SST model for the closure [32, 33]. The numerical schemes configurations and the set of constants

used for $k - \omega$ SST are reported in Appendix A.

3.1 Boundary conditions

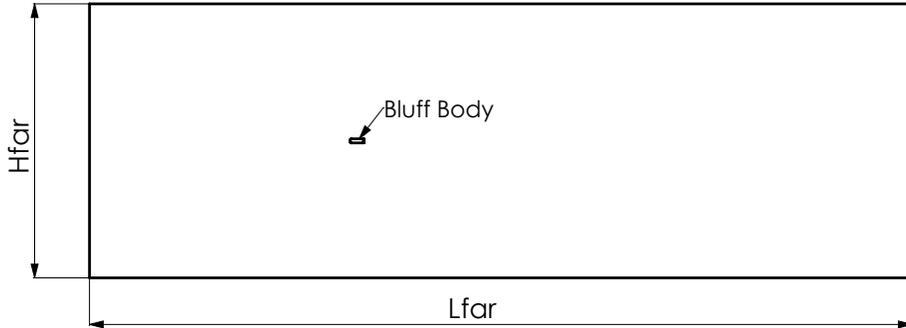


Figure 3: Geometry of the computational domain.

The external domain, shown in Fig. 3, is set large enough to minimize the effects of the artificial boundaries. It is 200 units long and 66 units high. The bluff body is positioned at half of the height and at $\frac{1}{3}$ of the length. While in the numerical experiments the velocity magnitude of the jets will vary according to the output of the policy network (see Section 4), for the assessment of grid and time convergence (see Section 3.3) it is fixed to a representative value of 2. A far-field boundary condition is prescribed on the outer box. The velocity U_∞ is set equal to 1 and the gauge pressure P_∞ equal to 0. The turbulence boundary conditions on the outer box are prescribed by the viscous ratio $r = \frac{\mu_t}{\mu}$ and turbulence intensity $I = \sqrt{\frac{3}{2}k}$. We set $r = 10$ and $I = 0.05$, which are typical values for an average level of turbulence. A no-slip boundary condition is applied on the body surface.

3.2 Mesh

The mesh is hybrid: it is made of a structured region adjacent to the body, and an unstructured region elsewhere (Fig. 4-5). The structured region contains 40 layers with a growth ratio of 1.1. This ensures to well resolve the boundary layer and the eddy viscosity creation in the front part. At the back, the number of layers is reduced to 20, because the triangular mesh is already fine enough and there is no advantage in increasing the number of layers. Along the curved edge (Fig. 6), there are no structured layers since the mesh is sufficiently fine to capture the interaction between the jet and the flow field. This implies that y^+ reaches a small enough

value by using triangular elements. Moreover, if rectangular elements had been used, the high diagonal flow crossing the elements would have compromised the convergence speed of the resolution.

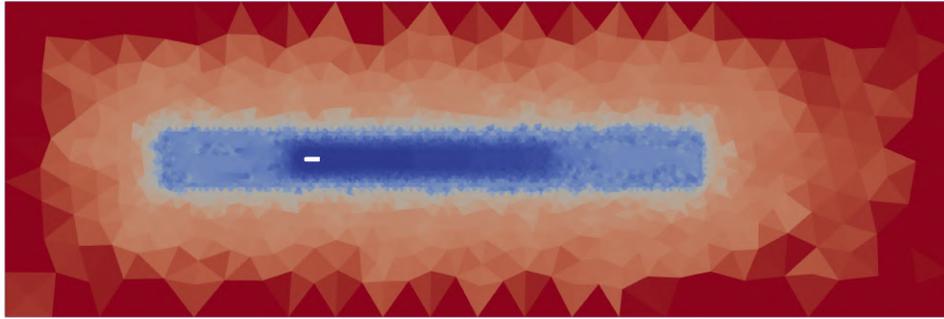


Figure 4: Qualitative representation of the grid cells size. The colors are proportional to the cell area.

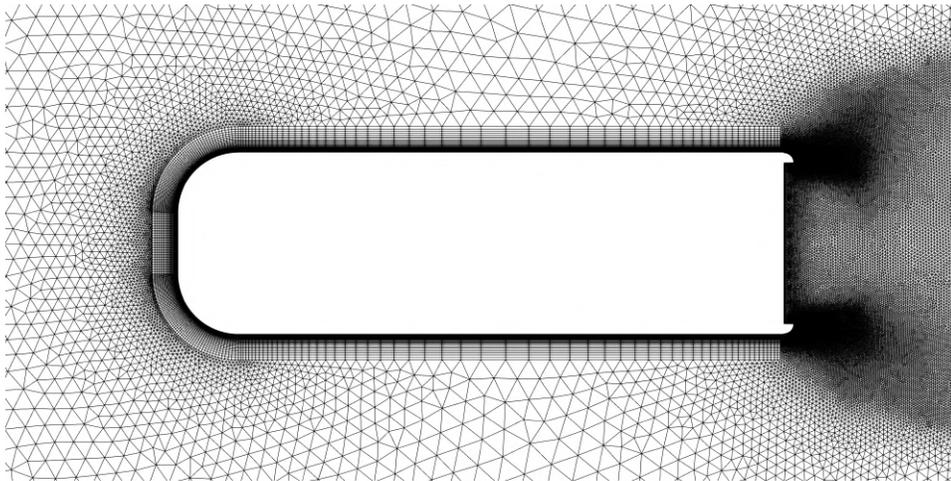


Figure 5: Mesh with 1.48×10^5 elements: Bluff body detail.

3.3 Grid and time convergence

The grid and time convergence (or independence) analysis is an important assessment to evaluate the characteristic size of a grid element and the size of the time step. It is carried out as follows: a coarse mesh and a large time step are initially considered and they are subsequently refined. This process continues until both

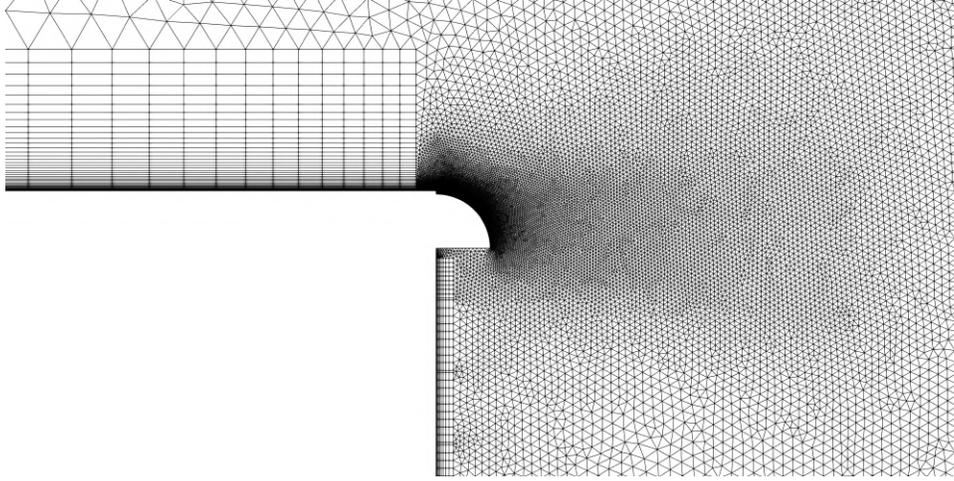


Figure 6: Mesh with 1.48e5 elements: Curved edge detail.

grid and time convergence is achieved. Table 1 shows the drag coefficient resulting from simulations with different grids and time steps. The table should be read from the top-left corner towards the bottom-right one where the grid element characteristic size and time step decrease and, conversely, the total number of the mesh elements increases. We can observe that the convergence for both parameters is reached with a time step $\Delta t = 2.5e-4$ and a mesh of $9.44e5$ elements. However, the computational complexity of the CFD simulation with the parameters obtained by the convergence analysis is incompatible with the reinforcement learning training. For this reason, we consider a mesh with $1.48e5$ elements and a time step of $4e-3$, adequate for the training.

Table 1: Time and grid convergence for the bluff body.

# elements	3.9e4	1.48e5	5.81e5	9.44e5
$\Delta t = 8e-3$	0.9145	-	-	-
$\Delta t = 4e-3$	0.9219	0.9749	-	-
$\Delta t = 2e-3$	0.9242	0.9785	-	-
$\Delta t = 1e-3$	-	0.9802	1.035	-
$\Delta t = 5e-4$	-	0.9811	1.038	1.045
$\Delta t = 2.5e-4$	-	-	-	1.045

4 Reinforcement Learning

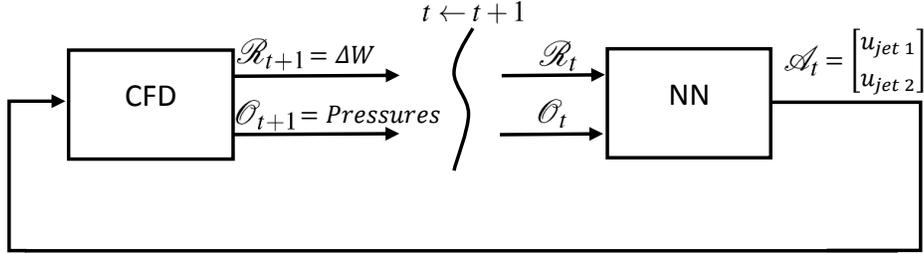


Figure 7: Agent-environment interaction. The left block represents the environment, the right one the agent. The environment is perceived through its observations, \mathcal{O}_t , that are measurements of the pressure on the rear surface at each time step. The reward is set equal to the power saving, ΔW . The action, \mathcal{A}_t , is a vector whose components are the instantaneous velocities of the jets.

We model the control of the flow rate of the nozzles as a partially observable Markov decision process $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{O}, \mathcal{T}, \mathcal{R}, \gamma \rangle$. Fig. 7 schematizes the elements of the Markov decision process specified to our problem, which are now described. Of the full state \mathcal{S} , corresponding to the whole fluid dynamics simulation, the agent has access to the observation \mathcal{O} , which consists of 12 pressure measurements at the back of the body (Fig. 8). Each pressure measurement is averaged over the time between two iterations, to filter out the high-frequency fluctuations. The pressure measurements persist in the observed state for 2 consecutive iterations, so that the agent has access to a the current and the previous observation (for a total of 24 values). This gives the agent access to an approximation of the first derivative of the pressure, necessary to understand whether a vortex is in the formation or removal phase. The transition dynamics $\mathcal{T} = P(s_{t+1}|s_t, a_t)$ is given by the evolution of the fluid dynamics simulation for $n = 50$ simulation time steps. The value n corresponds to approximately 1/5 unit of time. The fluid dynamics phenomena show an oscillating behaviour. Therefore, we can consider the lift coefficient, CL , and defining its period, called C_L -period. The C_L -period is about 4 units of time, so there are roughly 20 transitions per C_L -period, thus there are 20 actions and rewards per C_L -period. The value n is chosen as a compromise between temporal resolution and environment adaptation time. A higher value would imply weak time resolution, thus a coarser control. On the other hand, lower values would not be adequate, because the global phenomena, and thus the reward, are quite insensitive to high-frequency actions, making the credit assignment problem more difficult. The action $\mathcal{A} = (u_{jet_1}, u_{jet_2}) \in [0, 4] \times [0, 4]$ is the velocity of the

fluid emitted by the nozzle for the following transition. The parameter γ is the discount factor. The reward \mathcal{R} is the net power saving, namely the difference between the power of the drag of clean bluff body (see Section 3), W^{clean} , and the power of the drag of bluff body with controlled jets, $W^{CD+jets}$:

$$\mathcal{R}_{t+1} = W^{clean} - W_t^{CD+jets} = W^{clean} - W_t^{CD} - W_t^{jets}, \quad (1)$$

with:

$$W_t^{CD} = drag \cdot U_\infty = \frac{1}{2} \rho U_\infty^2 \left(\frac{1}{\Delta T} \int_t^{t+\Delta T} C_D \right) S U_\infty, \quad (2)$$

$$W^{clean} = drag_{clean} \cdot U_\infty = \frac{1}{2} \rho U_\infty^2 C_D^{clean} S U_\infty, \quad (3)$$

$$W_t^{jets} = \sum_{i=1,2} \frac{1}{\Delta T} \int_t^{t+\Delta T} \int_{slit} \left(\frac{1}{2} \rho |\mathbf{u}_{jet_i}|^2 + P_i \right) \mathbf{u}_{jet_i} \cdot \hat{\mathbf{n}}, \quad (4)$$

where S is the surface of the section of the bluff body. C_D consists of the pressure and viscous contributions, and the small thrust due to the mass flow exiting the body. For simplicity, we neglect all possible efficiency factors in the power computation. The pressure, P_i , in Eq. (4) is a gauge pressure referred to the pressure at infinity. The time average is made over one transition, so ΔT corresponds to the previously mentioned n time steps.



Figure 8: Position and numbering of pressure taps.

4.1 Experiments solution and algorithm settings

We use the Proximal Policy Optimization (PPO) algorithm [50]. The main hyperparameters are shown in Appendix C.

The actor policy and the critic are represented by two separate fully-connected neural networks, with 2 hidden layers of 30 units each. The policy network outputs the 2 action values, corresponding to the flow rate of the two jets, while the critic network outputs a single scalar, corresponding to the expected cumulative reward associated with the current state. The network size is selected in accordance to the theorem in [31, 40]. The total number of parameters is 1804 (actor) + 1711 (critic) for a total equal to 3515. The activation function, σ , is the Rectified Linear Unit (ReLU) function. Due to the computational complexity of the CFD simulation and the considerable number of time steps required to train a policy with reinforcement learning, the tuning of the hyperparameters of PPO and of the network architecture is beyond the scope of this work.

To challenge the computational complexity of the combined CFD simulation and reinforcement learning, the training phase is carried out in two steps. In the first step, the CFD simulation is run on a coarse mesh. This approximate simulation retains most of the properties of the complete one, while remaining computationally inexpensive. Indeed, by using a mesh with $2e4$ elements, we reduce the computation time by 92%, at the cost of an error less than 10% on the computation of the drag. As the control policy of the nozzles is initialized with random weights, it needs many interactions with the environment before finding an adequate control strategy. Once the performance of the agent saturates in this approximate environment, the second step consists in fine-tuning it in the CFD simulation with the full mesh.

5 Results

We now show the main results of the whole simulation process.

5.1 Policy learning

The graphs in Fig. 9 show the learning curve of the reinforcement learning agent in the two training steps described in Section 4.1. We can observe that, in the first training step, the agent starts the power consumption of the baseline, that is the clean bluff body represented in Fig. 2, and by accumulating experience it largely outperforms it. After approximately 170 training iterations, corresponding to 24 hours of CFD simulation, the reward reaches a plateau.

The plot of the learning curve in the second step of the training, performed on the finer mesh, shows that the impact of the fine tuning on the drag reduction is rather limited. While after approximately 60 training iteration, that corresponds approximately to 300 hours of simulations, the drag is further reduced by only 5%,

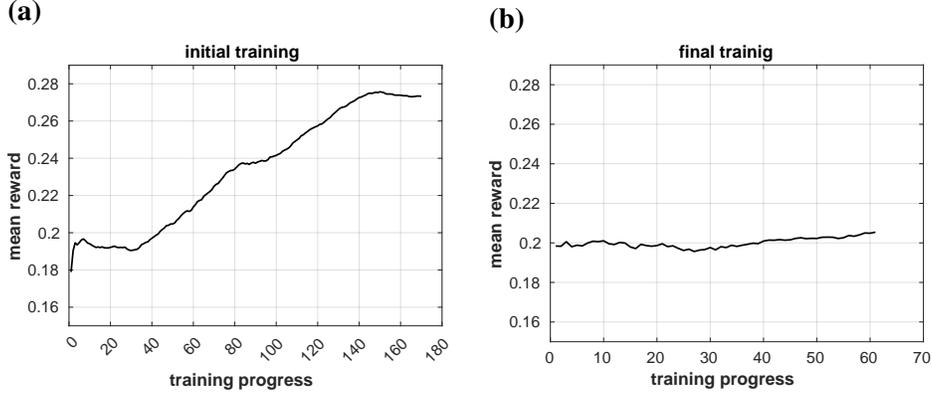


Figure 9: Training progress. **(a)** Initial training with a coarse mesh used to quickly adapt the neural network to the environment. The graph represents the mean rewards averaged over the N workers and over all the previous episodes versus the training progress. **(b)** Final training with a finer mesh. The gap between the mean rewards of the two phases is due to an improvement of the CFD accuracy, so a better reward calculation.

so most of the power saving was already achieved during the first training step. We highlight the fact that the mean rewards of the two training steps are different because of the different simulation environments. In fact, as per our convergence study on the mesh, the drag, and consequently the power of the drag force, computed in the coarse mesh is lower than the drag of the finer mesh. We recall that the initial training has the only aim to quickly adapt the neural network to the environment, saving a large amount of computation time.

The marginal drag reduction due to the final training phase indicates that the neural network obtained after the initial training can be successfully used without any adaptation, boasting similar performance to the adapted network. Already after the first training phase the two jets can interact with the vortices and obtain a consistent drag reduction even though the pressures on the rear surface for which it was trained are coarsely simulated, suggesting a good transfer learning property. The trained neural network is then applied to a simulation to study the learned control for the jets and to evaluate the power saving. The instantaneous powers W^{C_D} , W^{jet_1} , W^{jet_2} , and W^{C_D+jets} are reported in Fig. 10. With jet_1 and jet_2 , we denote the upper and the lower jet, respectively, and with the word *jets* we indicate the sum of the two jets, as in Eq. (4). Table 2 contains the averaged powers, evaluated from the aforementioned simulation, compared to the powers coming from a reference control with constant jets flow rate and velocities. The total power of the

bluff body with controlled jets is about 40% less than the clean bluff body power and almost 20% less than the constant jet case.

Table 2: Power saving and comparison with a constant-velocity jet case. The first row having $u_{jets} = 0$ corresponds to the clean bluff body. The power consumption of the controlled jet is comparable to the one of the constant-velocity jet equal to 2, but the C_D reduction is much higher, causing a net power saving of 40%.

u_{jets}	C_D	W^{C_D}	W^{jets}	W^{C_D+jets}	net saving %
0	1.18	0.589	0	0.589	0
1	1.15	0.577	0.001	0.578	2
2	1.01	0.506	0.020	0.526	11
3	0.83	0.414	0.076	0.490	17
4	0.57	0.283	0.183	0.466	21
5	0.42	0.209	0.364	0.573	3
controlled	0.67	0.323	0.028	0.356	40^{+11}_{-12}

According to our interpretation, the dynamics governing the jets-vortexes interaction is the following. The vortexes have low pressure in the core and, in the clean experiment are close to the rear surface as shown in Fig. 11 and Fig. 12. Essentially, the jets have two main effects: firstly, they push downstream the vortexes creation, promoting their rapid motion downstream, and leaving as a consequence a higher pressure near the wall. Indeed, Fig. 11(b) shows the distance from the wall at which the vortex is created and the high-pressure inter-space between them. Secondly, they reduce the vortexes intensity, as can be inferred from Figs 11 - 15.

We shall consider the Figs 12, which are few frames taken inside a period, looking at the vorticity to better appreciate the phenomena. Before the lower vortex is fully developed, the upper jet turns on, thus lowering the potential flow and pushing away the vortex in incipient creation. After its activation, an upper vortex arises and the lower jet turns on. Towards the end of the cycle, there are two vortexes not distinctly separated along the x coordinate, but after few instants, the positive vortex moves away vanishing and the negative one takes place recreating the initial situation. Moreover, this mechanism reduces the vertical oscillations of the wake (similar results were obtained in [12]) and makes the vortexes less intense and more irregular. Even the Strouhal number of vortex shedding is affected, decreasing from 0.24 to 0.23.

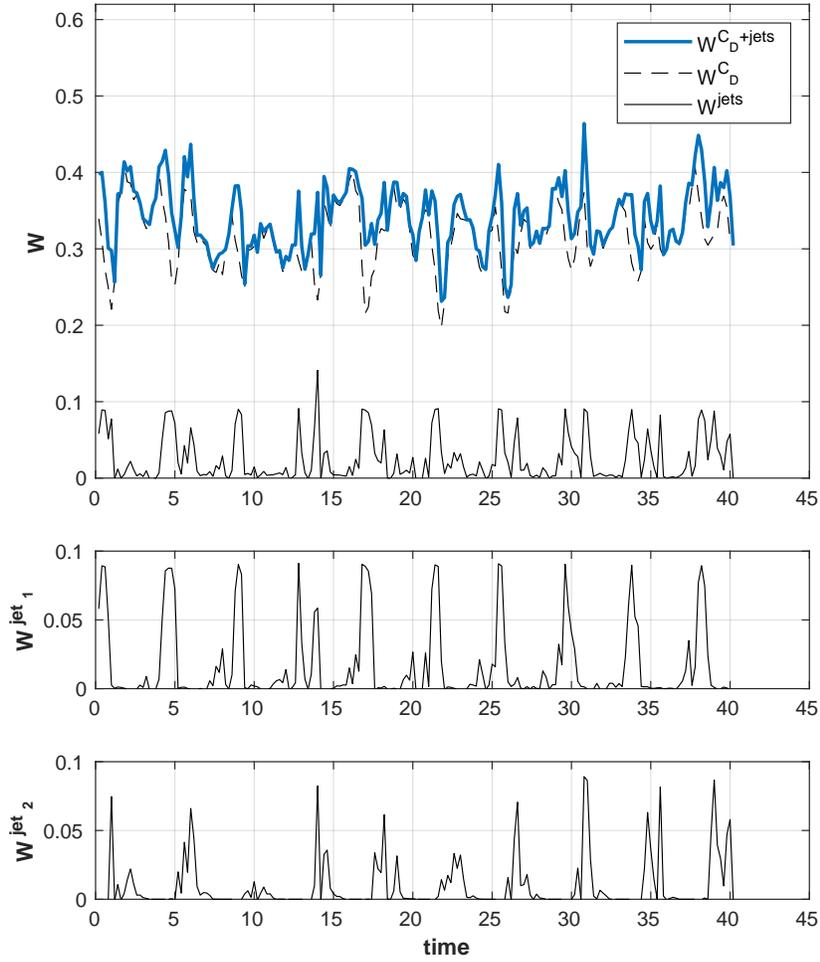


Figure 10: Upper panel: powers $W^{C_D+jets} = W^{C_D} + W^{jets}$, W^{C_D} , of the clean bluff body, and $W^{jets} = W^{jet_1} + W^{jet_2}$. Lower panels: powers W^{jet_1} , W^{jet_2} , of the upper and lower jet, respectively. The time laps is about $10 C_L$ -periods.

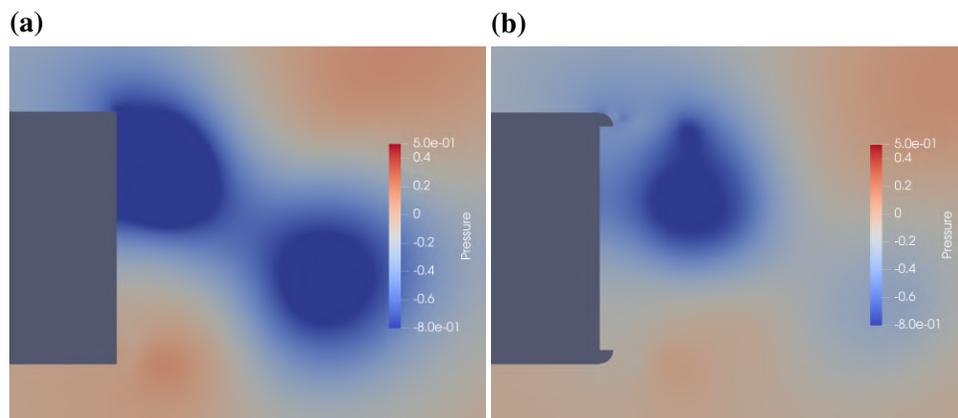


Figure 11: Snapshots of pressure distribution behind the bluff bodies. **(a)** Bluff body clean. The vortex is created close to the wall generating a notable low-pressure region. **(b)** Bluff body with controlled jets. The vortex is created far from the wall.

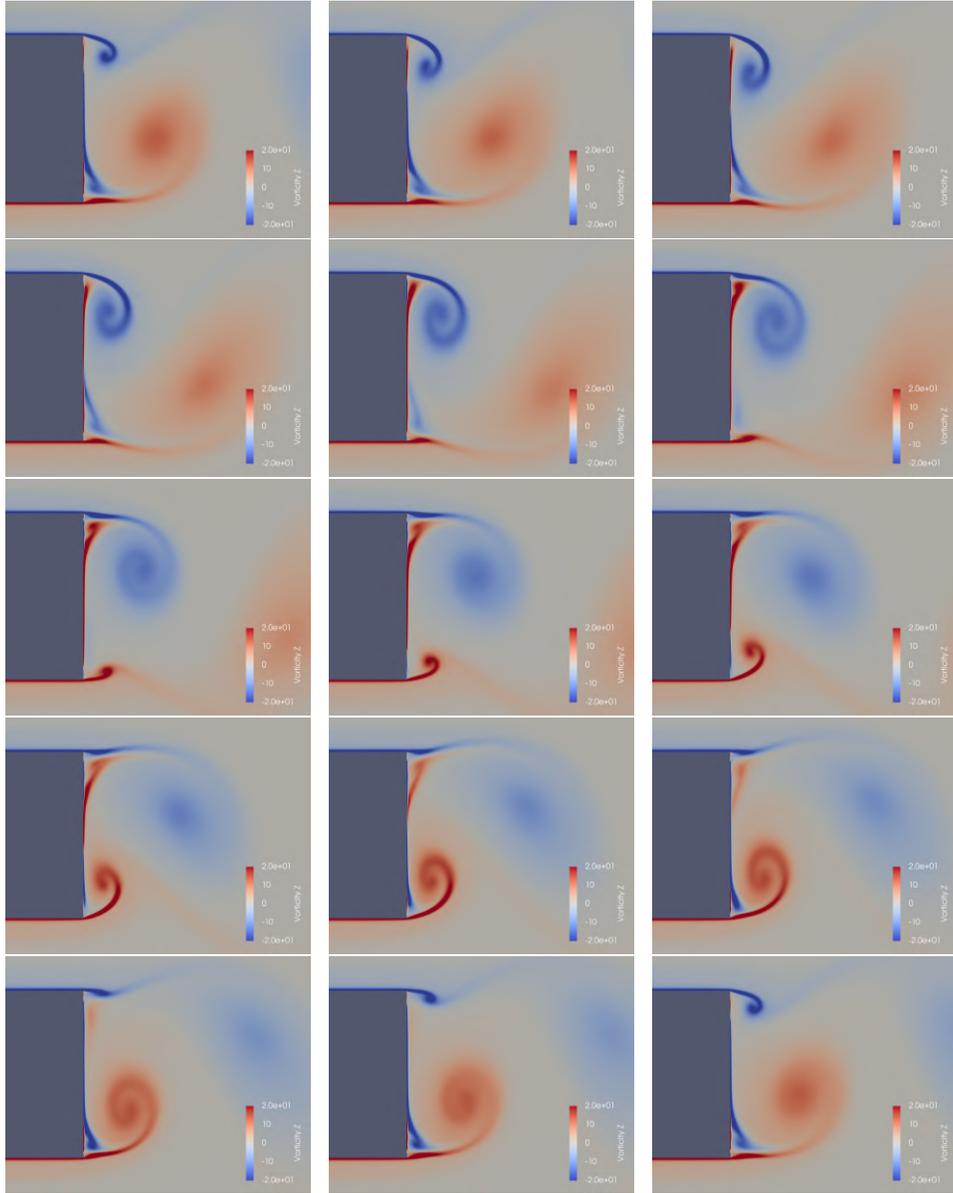


Figure 12: Time series of vorticity behind the body. The vortex formation and detachment is regular and periodic.

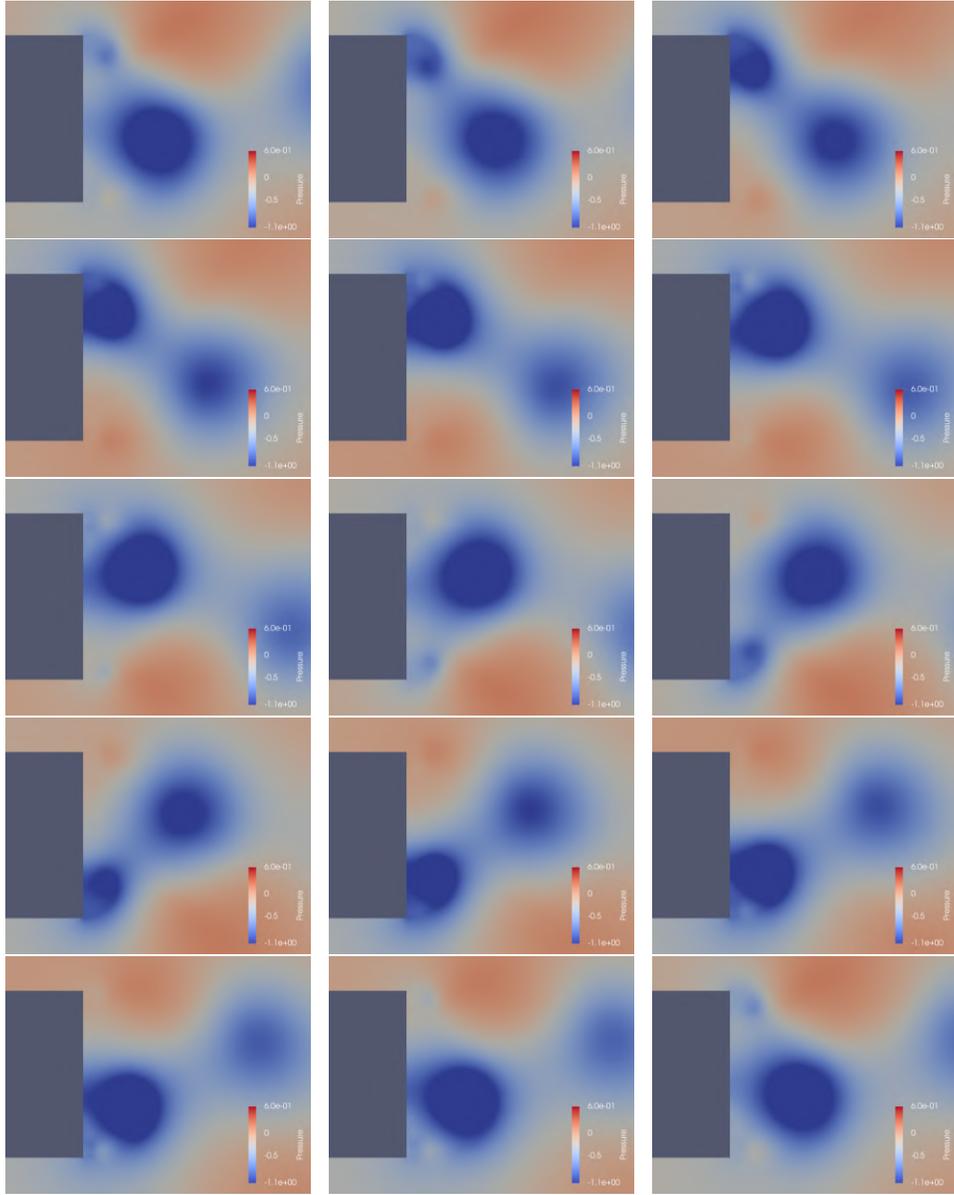


Figure 13: Time series of pressure behind the body. Low pressure regions caused by the cores of the vortices touches the rear surface of the body.

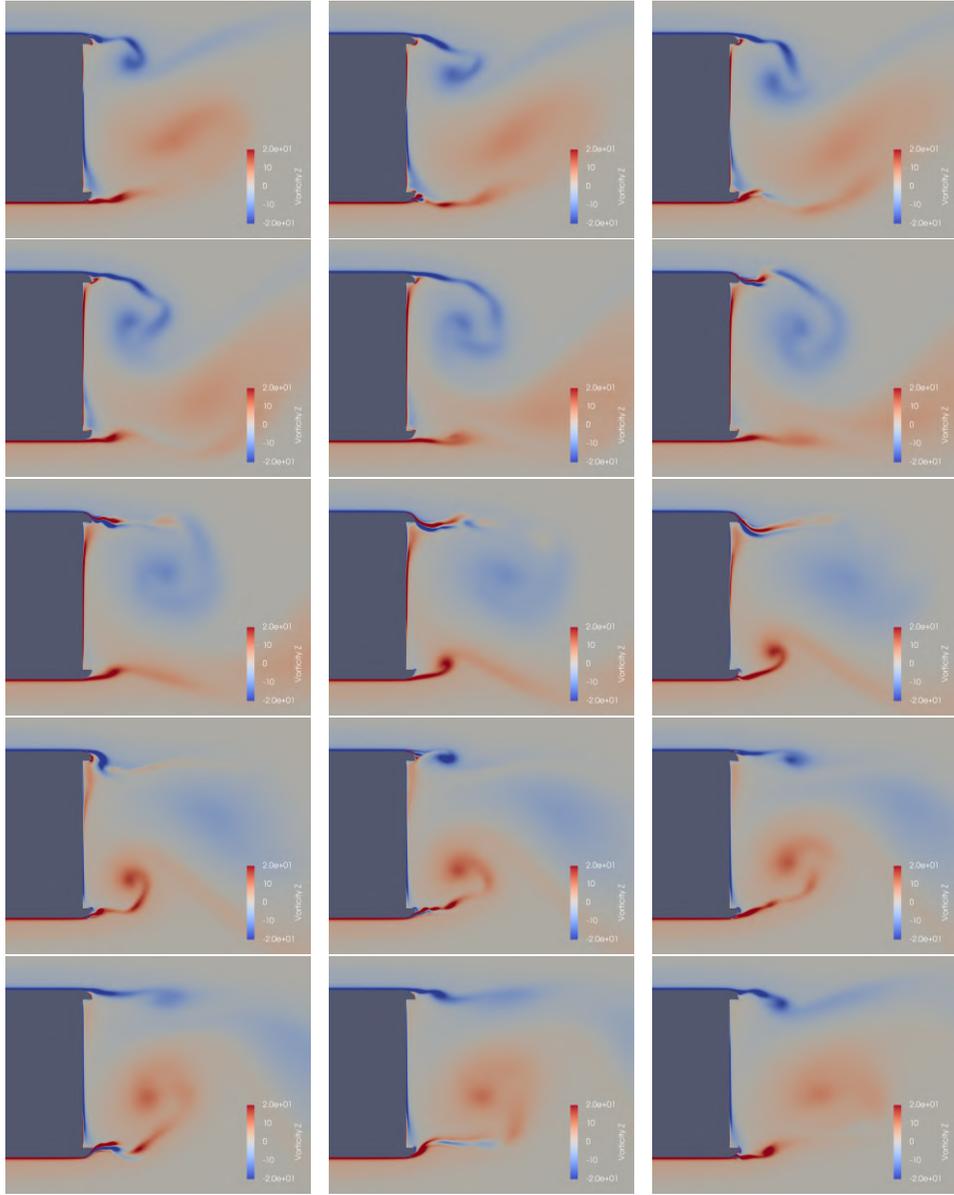


Figure 14: Time series of vorticity behind the body. The vortexes are strongly influenced by the fluid injection.

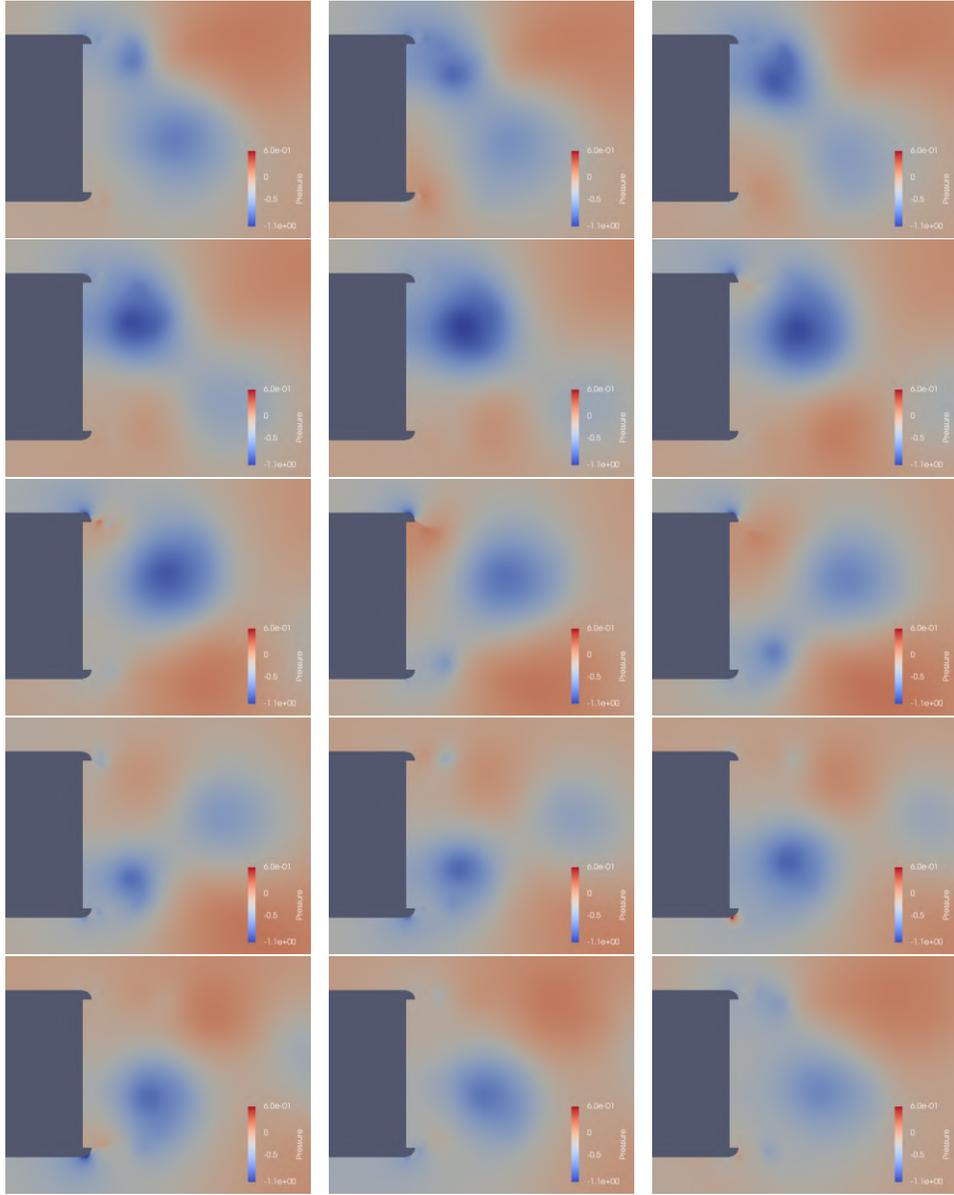


Figure 15: Time series of pressure behind the body. Thanks to the fluid injection, the vortices cores are weaker and further than in the clean bluff body.

6 Conclusions

In this work, we have applied reinforcement learning to the fluid dynamics problem of drag reduction of a bluff body. To achieve this, we focused on controlling two rear jets.

The reinforcement learning proved to be effective also in this fluid dynamics context. We have managed to train a neural network so that it interacts correctly with the environment and in particular with the wake vortices, despite the absence of a model of the environment but having only a poor vision of its response to the actions taken (data-driven approach). This interaction involves pushing downstream the vortex formation point and consequently removing the low-pressure region caused by the vortex core. In fact, the base pressure is mainly responsible for the drag reduction. Considering the overall system, i.e., the sum of the power of the drag force and the power expended by the controller, we have obtained a considerable saving of 40% compared to the uncontrolled case.

The results obtained amply justify the continuation of the research in this direction. Some research fields may be:

- Further improvement of the performances, with simultaneous optimization of geometry and control. In fact, during the jet-vortex interaction, the shape of the jet is also important, which is associated with the shape of the curved edge. Its optimization would improve the effectiveness of the jet. This simultaneous optimization should be performed within the training phase of the reinforcement learning, and the process could be enriched with the information deriving from the adjoint problem.
- In-depth study of the optimal control obtained and comparison of the results with other control methods to better define the potential of the reinforcement learning.
- Study on the memory of the neural network deepening the usefulness of recurrent neural network or multiple memory levels, i.e., saving also the antecedent states to the previous one.
- Test of control robustness and adaptation to other boundary conditions.
- Increased realism and accuracy of simulation and/or wind tunnel experiments, analyzing a more realistic geometry and possibly conforming to a practical case.

7 Acknowledgments

S. Micheletti thanks the PRIN research grant n. 20204LN5N5, *Advanced Polyhedral Discretisations of Heterogeneous PDEs for Multiphysics Problems* and the INdAM-GNCS 2022 Project, *Metodi di riduzione computazionale per le scienze applicate: focus su sistemi complessi*.

8 Conflict of interest statement

The authors have no conflicts to disclose.

A CFD settings

The constants used in the $k - \omega$ SST turbulence model, listed in Table 3, are found in [28].

Table 3: Constants used for the SST turbulence model.

coefficient	value
σ_{k1}	0.85
σ_{k2}	1
$\sigma_{\omega1}$	0.5
$\sigma_{\omega2}$	0.856
β_1	0.075
β_2	0.0828
γ_1	$\frac{\beta_1}{C_\mu} - \frac{\sigma_{\omega1}\kappa^2}{\sqrt{C_\mu}}$
γ_2	$\frac{\beta_2}{C_\mu} - \frac{\sigma_{\omega2}\kappa^2}{\sqrt{C_\mu}}$
C_μ	0.09
κ	0.41
a_1	0.31

The fluid dynamics equations are solved with a Finite Volume Method (FVM) set with the following main characteristics:

- The convective flux [13] is evaluated using the Flux Difference Splitting method with a MUSCL [55] second order reconstruction.
- The turbulent convective flux [13] is solved without the second-order reconstruction because of strong instabilities.
- Venkatakrishnan slope limiter [56] is applied.
- The derivatives of the velocity field in the viscous flux [3, 37, 38, 13] are retrieved from the Green-Gauss theorem.
- The time integration is performed using the first-order implicit Euler scheme.
- The linear system that arises after spatial discretization [13] is solved through BiCGSTAB with ILU preconditioner. The accuracy is set such that the error on the meaningful quantities (such as the drag coefficient) are accurate to within the

fourth digit. Decreasing this value would imply an increase of the computation time without an appreciable variation in the results.

B Hardware

The following is a list of the hardware used:

- Gigat (MOX - Dipartimento di Matematica of Politecnico di Milano): Xeon E5-2640 v4 @2.4GHz, RAM 64GB per node, 20 cores per node, max 2 nodes
- Gigatlong (MOX - Dipartimento di Matematica of Politecnico di Milano): Xeon E5-4610 v2 @2.3GHz, RAM 256GB per node, 32 cores per node, max 1 node at the time of this project.

C Reinforcement Learning settings

The main reinforcement learning settings are listed below: - Number of rollout workers, that is the number of environment run in parallel, equal to 4. From each rollout worker, a trajectory fragment of 160 samples is selected to fill the train batch of size $4 \times 160 = 640$. They are not ordered in a sequence but they are joint shuffled to avoid local overfitting.

- The number of epochs to execute per train batch is 30.
- The specific expression of objective function can be found in [50]. We use the Generalized Advantage Estimator (GAE) [49] as advantage.
- The learning rate is 0.0001 and the discount factor, γ , is 0.99.
- The initial coefficient for KL divergence is 0.2, while the target value for KL divergence is 0.01.

The size of the training dataset and the fragment length were object of study because they highly affect the training time. A small training dataset takes short time to be filled, while a big one requires computing more samples. From this aspect, smaller dataset should be preferred but, as confirmed by practice, they could not work well due to the physics of the problem: it becomes harder and harder to distinguish good actions from bad ones if the environment does not show their effects. This would lead to a poor estimation of the objective function gradient, and, consequently, to a bad weight update. The correct compromise between computation time and quality of the result is attained with a rollout worker fragment length of 160. The length of the trajectory fragment corresponds to, approximately, 8 C_L -periods. A shorter fragment length of, for example, 10 samples (0.5 C_L -periods)

makes the training impossible, whereas fragment length of, for example, 80 samples ($4 C_L$ -periods) is fine, but high rewards are not reached. Formally, our setup is episodic, with an episode maximum length of 160 samples. Despite the end of an episode, the environment is not reinitialized, so an episode starts from the last CFD-time steps of the previous episode. The episodes define only a cut in the continuous trajectory. This does not hold in case of numerical instability: if it happens, the CFD simulation is interrupted and relaunched from a given fixed initial condition. This is quite a rare event, so there is no manifestation of overfitting of the initial condition.

References

- [1] G. BEINTEMA, A. CORBETTA, L. BIFERALE, AND F. TOSCHI, *Controlling Rayleigh-Bénard convection via reinforcement learning*, *Journal of Turbulence*, 21 (2020), pp. 585–605.
- [2] L. BIFERALE, F. BONACCORSO, M. BUZZICOTTI, P. CLARK DI LEONI, AND K. GUSTAVSSON, *Zermelo’s problem: Optimal point-to-point navigation in 2D turbulent flows using reinforcement learning*, *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29 (2019), p. 103138.
- [3] J. BLAZEK, *Computational Fluid Dynamics: Principles and Applications*, Butterworth-Heinemann, Oxford, third ed., 2015.
- [4] E. BOHN, E. M. COATES, S. MOE, AND T. A. JOHANSEN, *Deep reinforcement learning attitude control of fixed-wing UAVs using proximal policy optimization*, 2019 International Conference on Unmanned Aircraft Systems (ICUAS), (2019), pp. 523–533.
- [5] R. D. BRACKSTON, J. M. GARCÍA DE LA CRUZ, A. WYNN, G. RIGAS, AND J. F. MORRISON, *Stochastic modelling and feedback control of bistability in a turbulent bluff body wake*, *Journal of Fluid Mechanics*, 802 (2016), pp. 726–749.
- [6] S. L. BRUNTON, B. R. NOACK, AND P. KOUMOUTSAKOS, *Machine learning for fluid mechanics*, *Annual Review of Fluid Mechanics*, 52 (2020), pp. 477–508.
- [7] M. A. BUCCI, O. SEMERARO, A. ALLAUZEN, G. WISNIEWSKI, L. CORDIER, AND L. MATHELIN, *Control of chaotic systems by deep reinforcement learning*, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 475 (2019), p. 20190351.

- [8] Z. CHEN AND C.-Y. WEN, *Flow control of a d-shaped bluff body using different dbd plasma actuators*, *Journal of Fluids and Structures*, 103 (2021), p. 103292.
- [9] H. CHOI, W. P. JEON, AND J. KIM, *Control of flow over a bluff body*, *Annual Review of Fluid Mechanics*, 40 (2008), pp. 113–139.
- [10] H. CHOI, J. LEE, AND H. PARK, *Aerodynamics of heavy vehicles*, *Annual Review of Fluid Mechanics*, 46 (2014), pp. 441–468.
- [11] M. CICOLIN, O. BUXTON, G. ASSI, AND P. BEARMAN, *The role of separation on the forces acting on a circular cylinder with a control rod*, *Journal of Fluid Mechanics*, 915 (2021), p. A33.
- [12] L. DALLA LONGA, A. S. MORGANS, AND J. A. DAHAN, *Reducing the pressure drag of a D-shaped bluff body using linear feedback control*, *Theoretical and Computational Fluid Dynamics*, 31 (2017), pp. 567–577.
- [13] T. D. ECONOMON, F. PALACIOS, S. R. COPELAND, T. W. LUKACZYK, AND J. J. ALONSO, *SU2: An open-source suite for multiphysics simulation and design*, *AIAA Journal*, 54 (2016), pp. 828–846.
- [14] L. C. EUN, A. S. M. RAFIE, S. WIRIADIDJAJA, AND O. F. MARZUKI, *An overview of passive and active drag reduction methods for bluff body of road vehicles*, *International Journal of Engineering and Technology(UAE)*, 7 (2018), pp. 53–56.
- [15] D. FAN, L. YANG, Z. WANG, M. S. TRIANTAFYLLOU, AND G. E. KARNIADAKIS, *Reinforcement learning for bluff body active flow control in experiments and simulations*, *Proceedings of the National Academy of Sciences*, 117 (2020), pp. 26091–26098.
- [16] A. M. FARAHMAND, S. NABI, AND D. N. NIKOVSKI, *Deep reinforcement learning for partial differential equation control*, in *2017 American Control Conference (ACC)*, 2017, pp. 3120–3127.
- [17] C. GARCÍA-BAENA, J. I. JIMÉNEZ-GONZÁLEZ, AND C. MARTÍNEZ-BAZÁN, *Drag reduction of a blunt body through reconfiguration of rear flexible plates*, *Physics of Fluids*, 33 (2021), p. 045102.
- [18] P. GARNIER, J. VIQUERAT, J. RABAULT, A. LARCHER, A. KUHNLE, AND E. HACHEM, *A review on deep reinforcement learning for fluid mechanics*, *arXiv:1908.04127*, (2021).

- [19] A. GATTO AND H. BABINSKY, *Investigation of passive porosity as a means for bluff-body drag reduction*, SAE Int. J. Commer. Veh., 11 (2018), pp. 65–73.
- [20] Y. HAFFNER, J. BORÉE, A. SPOHN, AND T. CASTELAIN, *Unsteady coanda effect and drag reduction for a turbulent wake*, Journal of Fluid Mechanics, 899 (2020), p. A36.
- [21] M. HASEGAWA AND H. SAKAUE, *Microfiber coating for drag reduction on a cylinder*, Journal of Fluids and Structures, 103 (2021), p. 103287.
- [22] L. HENNING AND R. KING, *Drag reduction by closed-loop control of a separated flow over a bluff body with a blunt trailing edge*, in Proceedings of the 44th IEEE Conference on Decision and Control, 2005, pp. 494–499.
- [23] T. HIRST, C. LI, Y. YANG, E. BRANDS, AND G. ZHA, *Bluff body drag reduction using passive flow control of jet boat tail*, SAE International Journal of Commercial Vehicles, 8 (2015), pp. 713–721.
- [24] K. KARTHIK, M. VISHNU, S. VENGADESAN, AND S. K. BHATTACHARYYA, *Optimization of bluff bodies for aerodynamic drag and sound reduction using CFD analysis*, Journal of Wind Engineering and Industrial Aerodynamics, 174 (2018), pp. 133–140.
- [25] M. KAZEMI, P. GHANOONI, M. MANI, AND M. SAEEDI, *Drag reduction of 3D bluff body using SDBD plasma actuators*, Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 235 (2021), pp. 1461–1480.
- [26] K. KLAUSMANN AND B. RUCK, *Drag reduction of circular cylinders by porous coating on the leeward side*, Journal of Fluid Mechanics, 813 (2017), pp. 382–411.
- [27] J. N. KUTZ, *Deep learning in fluid dynamics*, Journal of Fluid Mechanics, 814 (2017), pp. 1–4.
- [28] LANGLEY RESEARCH, *Turbulence modeling resource center*. <https://turbmodels.larc.nasa.gov/sst.html>.
- [29] M. F. A. LATIF, M. N. OTHMAN, Q. F. ZAHMANI, N. S. KHASHI'IE, B. E. ZHEN, M. F. ISMAIL, AND A. Y. ISMAIL, *Optimization of boundary layer separation reduction induced by the addition of a dimple grid on top of a bluff body*, Journal of Advanced Research in Fluid Mechanics and Thermal Sciences, (2019), pp. 173–182.

- [30] R. LI, B. R. NOACK, L. CORDIER, J. BORÉE, AND F. HARAMBAT, *Drag reduction of a car model by linear genetic programming control*, *Experiments in Fluids*, 58 (2017).
- [31] Z. LU, H. PU, F. WANG, Z. HU, AND L. WANG, *The expressive power of neural networks: A view from the width*, arXiv:1709.02540, (2017).
- [32] F. R. MENTER, *Two-equation eddy-viscosity turbulence models for engineering applications*, *AIAA Journal*, 32 (1994), pp. 1598–1605.
- [33] F. R. MENTER AND C. RUMSEY, *Assessment of two-equation turbulence models for transonic flows*, 1994. AIAA 1994-2343. Fluid Dynamics Conference.
- [34] S. M. NOS LANDIN, K. GHAZI-ZAHEDI, AND F. CICHOS, *Reinforcement learning of artificial microswimmers*, arXiv:1803.06425, (2018).
- [35] G. NOVATI, S. VERMA, D. ALEXEEV, D. ROSSINELLI, W. M. VAN REES, AND P. KOUMOUTSAKOS, *Synchronisation through learning for two self-propelled swimmers*, *Bioinspiration & Biomimetics*, 12 (2017), p. 036001.
- [36] A. R. OXLADE, J. F. MORRISON, A. QUBAIN, AND G. RIGAS, *High-frequency forcing of a turbulent axisymmetric wake*, *Journal of Fluid Mechanics*, 770 (2015), pp. 305–318.
- [37] F. PALACIOS, J. ALONSO, K. DURAISAMY, M. COLONNO, J. HICKEN, A. ARANAKE, A. CAMPOS, S. COPELAND, T. ECONOMON, A. LONKAR, T. LUKACZYK, AND T. TAYLOR, *Stanford University Unstructured (SU²): An open-source integrated computational environment for multi-physics simulation and design*, 2013. AIAA 2013-287. 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition.
- [38] F. PALACIOS, T. D. ECONOMON, A. ARANAKE, S. R. COPELAND, A. K. LONKAR, T. W. LUKACZYK, D. E. MANOSALVAS, K. R. NAIK, S. PADRON, B. TRACEY, A. VARIYAR, AND J. J. ALONSO, *Stanford University Unstructured (SU²): Analysis and Design Technology for Turbulent Flows*, 2014. AIAA 2014-0243. 52nd Aerospace Sciences Meeting.
- [39] H. PARK, D. LEE, W. P. JEON, S. HAHN, J. KIM, J. KIM, J. CHOI, AND H. CHOI, *Drag reduction in flow over a two-dimensional bluff body with a blunt trailing edge using a new passive device*, *Journal of Fluid Mechanics*, 563 (2006), pp. 389–414.

- [40] S. PARK, C. YUN, J. LEE, AND J. SHIN, *Minimum width for universal approximation*, arXiv:2006.08859, (2020).
- [41] A. PINELLI, M. OMIDYEGANEH, C. BRÜCKER, A. REVELL, A. SARKAR, AND E. ALINOVI, *The PELskin project: part IV-control of bluff body wakes using hairy filaments*, *Meccanica*, 52 (2017), pp. 1503–1514.
- [42] S. B. POPE, *Turbulent Flows*, Cambridge University Press, Cambridge, 2000.
- [43] Z. X. QIAO, G. MINELLI, B. R. NOACK, S. KRAJNOVIČ, AND V. CHERNORAY, *Multi-frequency aerodynamic control of a yawed bluff body optimized with a genetic algorithm*, *Journal of Wind Engineering and Industrial Aerodynamics*, 212 (2021), p. 104600.
- [44] G. REDDY, A. CELANI, T. J. SEJNOWSKI, AND M. VERGASSOLA, *Learning to soar in turbulent environments*, *Proceedings of the National Academy of Sciences*, 113 (2016), pp. E4877–E4884.
- [45] F. REN, H. HU, AND H. TANG, *Active flow control using machine learning: A brief review*, *Journal of Hydrodynamics*, 32 (2020), pp. 247–253.
- [46] F. REN, J. RABULT, AND H. TANG, *Applying deep reinforcement learning to active flow control in weakly turbulent conditions*, *Physics of Fluids*, 33 (2021), p. 037121.
- [47] I. SADREHAGHIGHI, *Artificial intelligence (AI) and deep learning for CFD*, tech. rep., CFD Open Series, 2021.
- [48] L. SALATI, *Aerodynamics of heavy trucks*. <http://hdl.handle.net/10589/138944>, 2018. Ph.D. Thesis.
- [49] J. SCHULMAN, P. MORITZ, S. LEVINE, M. JORDAN, AND P. ABBEEL, *High-dimensional continuous control using generalized advantage estimation*, arXiv:1506.02438, (2018).
- [50] J. SCHULMAN, F. WOLSKI, P. DHARIWAL, A. RADFORD, AND O. KLIMOV, *Proximal policy optimization algorithms*, arXiv:1707.06347, (2017).
- [51] N. A. SIDDIQUI AND M. A. CHAAB, *A simple passive device for the drag reduction of an ahmed body*, *Journal of Applied Fluid Mechanics*, 14 (2020), pp. 147–164.

- [52] K. SON, J. CHOI, W. P. JEON, AND H. CHOI, *Mechanism of drag reduction by a surface trip wire on a sphere*, Journal of Fluid Mechanics, 672 (2011), pp. 411–427.
- [53] R. SUTTON AND A. BARTO, *Reinforcement Learning*, The MIT Press, London, second ed., 2018.
- [54] N. TOUNSI, R. MESTIRI, L. KEIRSBULCK, H. OUALLI, S. HANCHI, AND F. ALOUI, *Experimental study of flow control on bluff body using piezoelectric actuators*, Journal of Applied Fluid Mechanics, 9 (2016), pp. 827–838.
- [55] B. VAN LEER, *Towards the ultimate conservative difference scheme. V. a second-order sequel to godunov’s method*, Journal of Computational Physics, 32 (1979), pp. 101–136.
- [56] V. VENKATAKRISHNAN, *On the accuracy of limiters and convergence to steady state solutions*, 1993. AIAA 1993-880. 31st Aerospace Sciences Meeting.
- [57] J. VIQUERAT, J. RABAULT, A. KUHNLE, H. GHRAIEB, A. LARCHER, AND E. HACHEM, *Direct shape optimization through deep reinforcement learning*, arXiv:1908.09885, (2020).
- [58] B. WANG, Z. YANG, AND H. ZHU, *Active flow control on the 25°Ahmed body using a new unsteady jet*, International Journal of Heat and Fluid Flow, 79 (2019), p. 108459.
- [59] Y. WANG, Z. SHEN, Z. LONG, AND B. DONG, *Learning to discretize: Solving 1D scalar conservation laws via deep reinforcement learning*, arXiv:1905.11079, (2020).
- [60] B. F. ZHANG, K. LIU, Y. ZHOU, S. TO, AND J. Y. TU, *Active drag reduction of a high-drag ahmed body based on steady blowing*, Journal of Fluid Mechanics, 856 (2018), pp. 351–396.
- [61] C. ZHENG, T. JI, F. XIE, X. ZHANG, H. ZHENG, AND Y. ZHENG, *From active learning to deep reinforcement learning: Intelligent active flow control in suppressing vortex-induced vibration*, Physics of Fluids, 33 (2021), p. 063607.

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 39/2023** Riccobelli, D.; Al-Terke, H. H.; Laaksonen, P.; Metrangolo, P.; Paananen, A.; Ras, R. H. A.; Ciarletta, P.; Vella, D.
Flattened and wrinkled encapsulated droplets: Shape-morphing induced by gravity and evaporation
- 38/2023** Africa, P. C.; Fumagalli, I.; Bucelli, M.; Zingaro, A.; Dede', L.; Quarteroni, A.
lifex-cfd: an open-source computational fluid dynamics solver for cardiovascular applications
- 37/2023** Regazzoni, F.; Pagani, S.; Salvador, M.; Dede', L.; Quarteroni, A.
Latent Dynamics Networks (LDNets): learning the intrinsic dynamics of spatio-temporal processes
- 34/2023** Caldana, M.; Antonietti, P. F.; Dede', L.
A Deep Learning algorithm to accelerate Algebraic Multigrid methods in Finite Element solvers of 3D elliptic PDEs
- 35/2023** Ferrari, L.; Manzi, G.; Micheletti, A.; Nicolussi, F.; Salini, S.
Pandemic Data Quality Modelling: A Bayesian Approach
- 32/2023** Gambarini, M.; Ciaramella, G.; Miglio, E.; Vanzan, T.
Robust optimization of control parameters for WEC arrays using stochastic methods
- 31/2023** Orlando, G.
Assessing ChatGPT for coding finite element methods
- 30/2023** Antonietti, P. F.; Bonizzoni, F.; Verani, M.
A DG-VEM method for the dissipative wave equation
- 29/2023** Carbonaro, D.; Mezzadri, F.; Ferro, N.; De Nisco, G.; Audenino, A.L.; Gallo, D.; Chiastra, C.; Morbiducci, U.; Perotto, S.
Design of innovative self-expandable femoral stents using inverse homogenization topology optimization
- 28/2023** Zingaro, A.; Vergara, C.; Dede', L.; Regazzoni, F.; Quarteroni, A.
A comprehensive mathematical model for myocardial perfusion