POLITECNICO DI TORINO
Repository ISTITUZIONALE

Deep reinforcement learning for active control of a three-dimensional bluff body wake

Original
Deep reinforcement learning for active control of a three-dimensional bluff body wake / Amico, E., Cafiero, G., Iuso, G.. - In: PHYSICS OF FLUIDS. - ISSN 1070-6631. - 34:10(2022), p. 105126. [10.1063/5.0108387]

Availability:
This version is available at: 11583/2976848 since: 2023-03-13T10:47:13Z

Publisher:
AIP Publishing

Published
DOI:10.1063/5.0108387

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

28 November 2026

Physics of Fluids

Deep reinforcement learning for active control of a three-dimensional bluff body wake

Cite as: Phys. Fluids **34**, 105126 (2022); https://doi.org/10.1063/5.0108387
Submitted: 08 July 2022 • Accepted: 16 September 2022 • Accepted Manuscript Online: 18 September 2022 • Published Online: 24 October 2022

E. Amico, G. Cafiero and G. Iuso

View Online Export Citation CrossMark

ARTICLES YOU MAY BE INTERESTED IN

DRLinFluids: An open-source Python platform of coupling deep reinforcement learning and OpenFOAM
Physics of Fluids **34**, 081801 (2022); https://doi.org/10.1063/5.0103113

From active learning to deep reinforcement learning: Intelligent active flow control in suppressing vortex-induced vibration
Physics of Fluids **33**, 063607 (2021); https://doi.org/10.1063/5.0052524

Active control for enhancing vortex induced vibration of a circular cylinder based on deep reinforcement learning
Physics of Fluids **33**, 103604 (2021); https://doi.org/10.1063/5.0063988



AIP Publishing

© 2022 Author(s).

Physics of Fluids ARTICLE scitation.org/journal/phf

Deep reinforcement learning for active control of a three-dimensional bluff body wake

Cite as: Phys. Fluids **34**, 105126 (2022); doi: 10.1063/5.0108387
Submitted: 8 July 2022 · Accepted: 16 September 2022 ·
Published Online: 24 October 2022

View Online Export Citation CrossMark

E. Amico, G. Cafiero,[a] and G. Iuso

AFFILIATIONS

Dipartimento di Ingegneria Meccanica e Aerospaziale, Corso Duca degli Abruzzi 24, Politecnico di Torino, 10129 Torino, Italy

[a]Author to whom correspondence should be addressed: gioacchino.cafiero@polito.it

ABSTRACT

The application of deep reinforcement learning (DRL) to train an agent capable of learning control laws for pulsed jets to manipulate the wake of a bluff body is presented and discussed. The work has been performed experimentally at a value of the Reynolds number $Re \sim 10^5$ adopting a single-step approach for the training of the agent. Two main aspects are targeted: first, the dimension of the state, allowing us to draw conclusions on its effect on the training of the neural network; second, the capability of the agent to learn optimal strategies aimed at maximizing more complex tasks identified with the reward. The agent is trained to learn strategies that minimize drag only or minimize drag while maximizing the power budget of the fluidic system. The results show that independently on the definition of the reward, the DRL learns forcing conditions that yield values of drag reduction that are as large as 10% when the reward is based on the drag minimization only. On the other hand, when also the power budget is accounted for, the agent learns forcing configurations that yield lower drag reduction (5%) but characterized by large values of the efficiency. A comparison between the natural and the forced conditions is carried out in terms of the pressure distribution across the model's base. The different structure of the wake that is obtained depending on the training of the agent suggests that the possible forcing configuration yielding similar values of the reward is local minima for the problem. This represents, to the authors' knowledge, the first application of a single-step DRL in an experimental framework at large values of the Reynolds number to control the wake of a three-dimensional bluff body.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0108387

I. INTRODUCTION

The effect of aerodynamic drag in the automotive sector covers a key role in fuel consumption, thus directly relating to hazardous gas emissions. At typical speeds on motorways, 130 km/h, the aerodynamic drag can account for 80% of the total drag; at 80 km/h, it represents about 50%.[1] According to a study conducted in the US and the UK, fuel consumption impacts 20%–30% on the operating costs.[2] The limited autonomy of the electric vehicles also represents a key driver for the improvement of the aerodynamic performance of heavy duty vehicles.

Although these data show the need to minimize the aerodynamic drag, heavy duty vehicles' design is generally driven by the need to accommodate goods, hence to maximize their storage. The shape is then quite often poor from the aerodynamic standpoint, being representative of a bluff body geometry, whose near wake is one of the key responsible for the high values of the drag.

Taming the dynamics of the wake is, therefore, a key enabler toward obtaining significant drag reductions. In this sense, flow control techniques can cover a key-role. They can be generally divided into two groups: passive[3,4] and active.[5–7] The former does not require an external power supply to function. However, their implementation can be cumbersome, especially due to restrictions related to certification issues and/or the presence of appendages that can increase the length of the vehicle. Some examples are vortex generators,[8] flaps,[9] automatic mobile flap,[10] and geometric modifications of the base.[11]

The active flow control can be open or closed-loop: in the latter case, the output of one or more sensors feeds back to the control system to decide the next actuation. Closed-loop flow control offers further degrees of freedom to improve actuation efficiency. It requires mathematical or conceptual models that link actuation effects and sensor information. The model and the controller have to be selected with respect to the behavior of the flow. Examples are found in the literature for the control of the wake of bluff bodies using low order modeling,[12] opposition control,[13] and linear control[14] to cite some examples. It is also worth noticing that similar approaches can be followed to provide flow field estimations from near wall sensors.[15,16]

The fast development of model-free and data-driven techniques over the last few years opens the path to a dramatically different approach. Genetic algorithms (GA) and artificial neural networks

Published under an exclusive license by AIP Publishing

(ANNs) are showing the most promising results in solving high dimensional and non-linear problems. There are several examples of the application of GA to Active Flow Control (AFC): Minelli *et al.*[5] and Amico *et al.*[17] determined the forcing conditions to minimize the drag of different types of bluff bodies, such as square prisms, D-shaped, and road vehicles. Zhou *et al.*[18] implemented the GA to manipulate the acoustic signature of turbulent jets. The bottleneck of these techniques is the learning time, along with the complexity of the task learned, which can hardly be linked to the underlying physics.

The growth of Machine Learning (ML) and artificial intelligence has given a strong impetus to artificial neural networks. Among others, the solutions that have shown the most promising results are obtained using ANN trained through Deep Reinforcement Learning (DRL[19]). The DRL models an agent that interacts with an environment (in this case, the flow field) through a series of actions (the control laws) obtaining a given reward (the function that needs to be designed and optimized). The final aim is to maximize this reward. The training of the ANN occurs through a series of trials and errors, and it is based on the reward obtained for a given action.

DRL showed astonishing results in training control laws in the case of 2D numerical simulations at low to moderate values of the Reynolds number.[20,21] A review of the recent applications of DRL to fluid mechanics can be found in Garnier *et al.*,[22] Pino *et al.*,[23] and Viquerat *et al.*[24]

The present investigation builds on the results obtained in Cerutti *et al.*,[25] where an open loop control of the wake of a bluff body through jets located at the base of the model was implemented. In this work, it is proposed a new framework based on artificial intelligence where the jets' actuation is defined through a single-step DRL. It is worth explicitly mentioning that the implementation of the DRL is instrumental in discovering flow control strategies without providing any information to the algorithm in addition to the representation of the flow. This approach disentangles the control strategy from the modeling of the wake, thus shifting the interest toward the interpretability of the actuation, which will be tackled in future investigations.

At the time of the submission of this manuscript, the applications of DRL were limited only to much smaller values of the Reynolds number and to numerical simulations. The only exception, to the authors' knowledge, is the case of Fan *et al.*[26] In that case, the control of a cylinder wake is performed through the rotation of two smaller cylinders, which are actuated at the same rotation speed. In the present investigation, a more complex flow, such as the one produced from a bluff body, and a more complex space in terms of the actions are experimentally investigated and controlled through DRL.

This paper is structured as follows: in Sec. II, the experimental setup and the measurement techniques are described; in Sec. III, a brief introduction to DRL with some key definitions is reported. In Sec. IV, the results of the training and evaluation of the agent are discussed. Finally, in Sec. V, the main conclusions are drawn.

II. EXPERIMENTAL SETUP

The experiments are carried out in the open-circuit wind tunnel at Politecnico di Torino. The flow is accelerated by two fans located upstream of a stagnation chamber, and it is then conveyed to the test section through a convergent. The test section has a rectangular cross section (0.9 m high and 1.2 m wide) and a length of 6.5 m; it is characterized by a small divergence angle of ∼1% to account for the growth of the boundary layer on the wind tunnel walls.

The test speed is measured by evaluating the pressure ratio between the inlet and the outlet of the convergent, which was carefully calibrated against the readings of a Pitot tube located in the test section.

The model used[25,27] is representative of a 1:10 scaled model of a square-back road vehicle typically employed as heavy duty vehicle. It is characterized by a section with a maximum height $H = 0.200$ m and a width $W = 0.170$ m, while its length $L = 0.412$ m. A back slant angle of 10° (as shown in Fig. 1) allows to emphasize the effect of the active control system as shown in Barros *et al.*,[28] exploiting the Coanda effect.

The model is held in position with a vertical strut connected to its top surface. The resulting gap between the underbody and the fixed wind tunnel floor is equal to $g = 20$ mm ($g/H = 0.1$). The strut is embedded in an aerodynamically shaped wing profile (NACA 0020, with maximum thickness $t/c = 0.2$ and $t/W = 0.07$), and it rigidly connects the model to a load cell located outside the test section to perform the drag measurements. The fairing is also used to carry the pneumatic lines that supply the air jets, the connection cables to the pressure scanner that is embedded within the model and the acquisition signals. The frontal area of the model is equal to 3.1% of the cross section of the wind tunnel; this value increases to 4.5% considering also the strut fairing.

The boundary layer growth developing on the wind tunnel floor has been minimized by implementing a suction slit at about two model lengths upstream of the model's nose. The resulting boundary layer characteristics, as measured by dedicated hot wire anemometry experiments, are such that $\delta^*/g = 0.07$, with δ^* being the displacement thickness of the boundary layer. This value is in good agreement with other investigations performed with a fixed floor.[29]

A. Flow control system

The flow control system is constituted of four air jets located along the edges of the model's base, as schematically represented in Fig. 1.

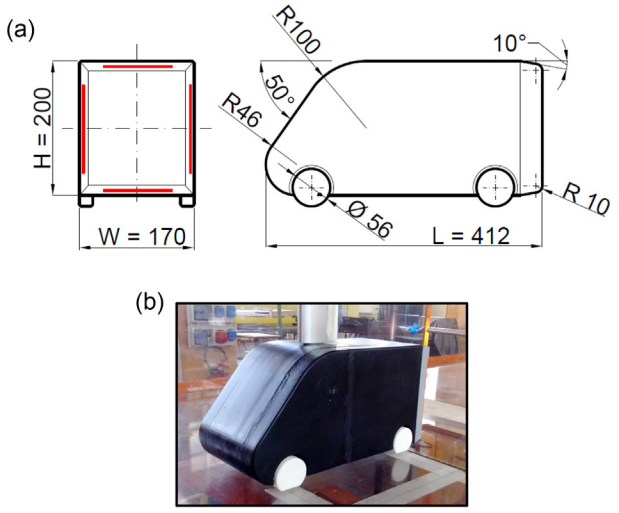


FIG. 1. (a) Schematic representation of the model used for the wind tunnel experiments; the continuous red lines represent the injection slots of the active flow control system. Units are in mm. (b) Picture of the model mounted in the wind tunnel.

Published under an exclusive license by AIP Publishing

The individual actuator is made up of a cylinder with a rectangular slot of 1 mm along its length, as the one implemented in previous investigations.[25] The geometry is selected to allow the orientation of the air jet along different injection angles. In the present case, a fixed injection angle of 65° toward the center of the wake is considered. The slots have a length of 104 mm in the case of the top and bottom slots, and 132 mm in the case of lateral slots. Hot wire anemometry measurements were performed to characterize the velocity profile at the exit section of the slot. The velocity profile was found to be uniform for at least 95% of the slot's length.

A schematic representation of the air supply that feeds the jets is reported in Fig. 2. A pneumatic line with a maximum operating pressure of 10 bar feeds the system. The flow rate that feeds each of the slots is measured using three independent flow meters (FM). The flow control system is designed in a way such that it allows the independent activation of the top, bottom, and lateral jets. It must be specifically noted that the lateral jets are connected to the same valve and, as such, will be actuated according to the same control law.

The flow rate is varied using three solenoid valves that can be modulated according to different control laws, such as continuous, square wave, sinusoidal wave, and sawtooth (V). In the present case, a sinusoidal wave is selected. The DRL algorithm will learn the control law in terms of the amplitude and the frequency for each one of the three different actuators (top, bottom, and the two lateral jets).

B. Measurement techniques

The model is instrumented with static pressure taps using multi-input pressure transducers and capacitive microphone capsules to sample the fluctuating pressure signals. In particular, the signals relating to static pressures are acquired using a Scanivalve connected to a Smart Zoc 100. The transducer has a full scale of ± 2.5 kPa and accuracy of 0.15%*FS*. The system allows the acquisition of 64 simultaneous channels at a sampling frequency $f_a = 50$ Hz. The 64 channels are connected by pneumatic tubes with an internal diameter equal to $\Phi = 1$ mm to the pressure taps. The pressure taps are distributed as follows: 31 populate the model's base, and 33 are distributed across the front, top, and lateral surfaces of the model.

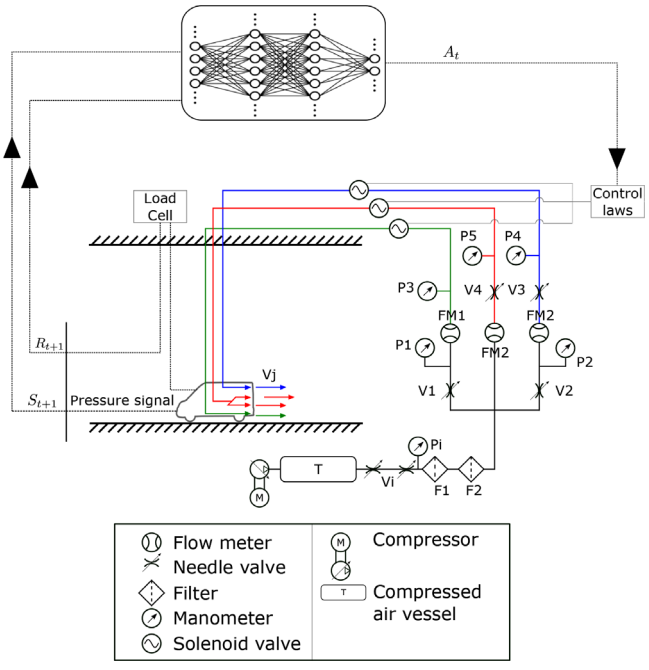


FIG. 2. Schematic representation of the implementation of the DRL algorithm, the pulsed jet actuators, and the control system.

Furthermore, 16 microphone sensors with an external diameter of 9.8 mm and a sensing element with a diameter of 1 mm were used to measure the fluctuating pressure signals. Twelve microphones are placed on the base, and the remaining four are placed on the four side faces at a distance of 10 mm from the edges of the base. The microphones are characterized by a flat response in the range of frequencies $0.005 - 13$ kHz and sensitivity of -60 ± 3 dB. All the microphone probes were calibrated using a *Bruel & Kjaer* probe, as reported in Ref. 30. Both for the measurements and the calibration, a *pinhole* configuration was used. The electrical signals were filtered to eliminate spurious contributions.[25]

The drag measurements were performed using a one-axis load cell *Dacell UU-K002* with a full scale $F.S. = 2\,Kg_f$ and a rated output equal to $1.5\,\mathrm{mV/V} \pm 1\%$. The excitation voltage was set to 10 V using a stabilized power supply AL862D 0–30 V. The output signal is then amplified using a dedicated conditioning module provided by the load cell maker.

The load cell and the microphone signals were simultaneously sampled using a NI-cDAQ chassis with dedicated A/D converter modules NI-9215, with resolution equal to 16 bits and a full scale equal to ± 10 V. The analog signal of the load cell is converted into drag through a calibration mapping.

Table I summarizes the sampling parameters that were selected, in regard to the static pressure, fluctuating pressure, and load cell measurements. The selected acquisition time was defined as a trade-off between the need to obtain well-converged statistics and the requirements dictated by the training of the DRL algorithm. In particular, the measurements were performed for at least 1000 timescales based on the freestream speed and characteristic length of the model.

III. DEEP REINFORCEMENT LEARNING

Deep Reinforcement Learning (DRL) belongs to the family of Machine Learning (ML) algorithms.

An agent, typically modeled with an artificial neural network (ANN), is trained to learn strategies to optimize a reward, by interacting and exchanging information with the environment. Figure 2 shows a schematic representation of the DRL.

In this work, a vanilla policy-gradient method implemented within the library TensorForce[31] has been selected as agent. This is a configurable library allowing to efficiently represent a broad range of

TABLE I. Summary of the sampling parameters.

Measurement	t_a (s)	f_a (Hz)
Static pressure	20	50
Fluctuating pressure	20	2000
Load cell	20	2000

Published under an exclusive license by AIP Publishing

agents for DRL. They act with a policy parameterized by a neural network and exploit memory modules to perform updates.

The hyperparameters selected for the present investigation are summarized in Table II. The weights of the neural network are updated every five timesteps, while the total number of timesteps included in one episode is equal to 15. The choice of updating the weights of the network every five timesteps is driven by the will of mitigating any possible spurious results associated with the experimental framework, which would hinder the convergence of the algorithm. This aspect can be further investigated in future works in order to provide a trade-off between the averaging time of the inputs to the network and the number of timesteps between updates of the weights of the network.

The horizon is set to one, which has been previously referred to as *degenerate DRL*[32] in problems of shape optimization, or it could be possibly close to the implementation indicated as single-step Proximal Policy Optimization (PPO)[33] implemented for heat-transfer problems and shape optimization.[34]

This approach was deemed acceptable as the objective is to determine the value of the actions yielding the optimization of the reward, starting from the uncontrolled condition. In other words, it is not of interest the sequence of the actions, but only the final value of the action yielding the optimized reward. In particular, the actuation can be seen as a direct mapping between the uncontrolled condition and the most profitable control condition. This approach is less general, yet it allows to avoid problems such as the difference in time scales between the experimental framework and the agent. It must be explicitly noted also that the single-step DRL is generally thought for a constant input case. In the present investigation, at the beginning of each episode, the network restarts from the uncontrolled condition, which will be constant in time to within the uncertainty of the experiment. The results reported in this manuscript show that the single-step approach is also applicable to cases where the state is not necessarily constant, provided that the variations are not too significant with respect to the target state. In that case, the algorithm would necessarily fail as the network would learn a mapping between a non-representative state and a controlled configuration.

The default parameters were selected to define the neural network, hence using two fully connected hidden layers with 64 elements each, plus the input layer of the observations and the output layer. The hyperbolic tangent was selected as the activation function.

An advantage of this method is related to the relatively fast convergence time, with respect to cases, for example, based on evolutionary algorithms.[17] It is worth evidencing that in the experimental environment, the time required for the convergence of the algorithm covers a key role in the problem definition. Particularly long convergence times imply issues related to the variation of the freestream conditions due to temperature drift, hysteresis phenomena, offset variation of the measurement system, and, last but not least, costs related to facilities.

The environment, in the present investigation, can be characterized by the static and fluctuating pressure signals acquired across the base of the model.

We introduce two different definitions of the state, agent A and agent B, to provide insights into the effect of the number of probes on the convergence of the algorithm (also summarized in Table III):

- Agent A: pressure coefficient Cp measured across the model's base. In this case, 29 static pressure signals are acquired at each timestep of the DRL. The pressure signals that define the state are averaged over the acquisition time (as reported in Sec. II), and then normalized according to the definition of the pressure coefficient

$$Cp = \frac{\bar{p} - \bar{p}_\infty}{0.5 \cdot \rho \cdot U_\infty^2}, \tag{1}$$

  where the subscript ∞ indicates the freestream conditions.
- Agent B: pressure coefficient Cp measured across the model's base in addition to the fluctuating pressure sensors. The 29 static pressure probes are complemented by 14 microphone probes that populate the model's base. The state is, therefore, complemented by the root mean square (rms) of the fluctuating pressure, calculated as

$$Cp_{RMS} = \frac{\sqrt{\overline{(p')^2}}}{0.5 \cdot \rho \cdot U_\infty^2}, \tag{2}$$

  where p' is the voltage output from each microphone sensor.

The actions are how the agent can interact with the environment. In this case, the actions are the parameters of the control laws. A sinusoidal control law was chosen for each of the three actuators, as schematically represented in Fig. 2.

The three control laws will have zero phase and a non-zero minimum value of the voltage. For an operating voltage equal to $V_{\min} = 5.5\,\mathrm{V}$, the mass flow rate would equal zero. Therefore, the parameters to be defined are the frequency (f_j) and the maximum operating voltage ($V_{j\max}$) of each actuator. Thus, the DRL algorithm will optimize a total of six parameters (three frequencies and three voltages), with the resulting control law

$$V_j = \frac{V_{jmax} - V_{\min}}{2} \cdot \sin(2\pi f_j \cdot t) + \frac{V_{jmax} + V_{\min}}{2}, \tag{3}$$

where j indicates the lateral, top, or bottom actuator. The resulting value of the voltage is linked to the exit speed of each jet through an *a priori* defined calibration mapping.

TABLE II. Agent hyperparameters.

Parameter	Value
Update (timesteps)	5
Horizon	1
Max episode timesteps	15
Entropy regularization	0.01

TABLE III. Summary of the states and the rewards explored in the present investigation.

	Agent A	Agent B
Case 1	State: Cp, Reward: ΔCd	State: $Cp + Cp_{RMS}$, Reward: ΔCd
Case 2	State: Cp, Reward: $\Delta Cd + k_\mu$	State: $Cp + Cp_{RMS}$, Reward: $\Delta Cd + k_\mu$

Published under an exclusive license by AIP Publishing

As suggested in the literature,[19,35] the actions (amplitude and frequency of the actuation) are normalized, so that the variation is within the range between −1 and 1. The normalization of the quantities takes place according to the following definitions:

$$A_e = 2 \cdot \frac{V_{jmax} - V_{\min}}{V_{\max} - V_{\min}} - 1, \tag{4}$$

where the maximum and minimum values of the voltage are 9.5 and 5.5 V, respectively,

$$A_f = 2 \cdot \frac{f_j - f_{\min}}{f_{\max} - f_{\min}} - 1, \tag{5}$$

and the maximum and minimum frequency are 30 Hz and 0, respectively. When each of the two parameters attains a value equal to −1, this corresponds to a null value of the voltage or frequency, while when it attains a value of 1, the voltage and the frequency are maximum. Furthermore, the condition $A_f = 0$ corresponds to the case of a steady jet injection.

A. Power budget

The application of an active flow control system to real problems must necessarily account for the power necessary to run the system.

This means that the drag reduction must be obtained efficiently. It is important to point out that the definition of efficiency cannot be obtained trivially, as suggested by Choi *et al.*[36]

A generally well-accepted definition, that is only related to the fluid dynamics benefit associated with the control system, is the one suggested by Englar[37] who analyzed the sensitivity of the wake control through the relationship between the variation of the drag coefficient and the jet momentum coefficient C_μ

$$\frac{\Delta Cd}{C_\mu} = \frac{D_0 - D}{\frac{1}{2}\sum_{j=1}^{N_{jets}} \rho A_j U_j^2}, \tag{6}$$

where D_0 indicates the drag in the natural case (i.e., without control), A_j is the cross-section of the jth jet, and V_j is the exit speed of the jth jet. The denominator term takes into account the amount of energy consumed. Equation (6) can take both positive and negative values; in particular, drag reducing configurations are associated with positive values. Furthermore, values of $\frac{\Delta C_d}{C_\mu} > 1$ correspond to energy-efficient configurations.

The last term missing to fully define the problem is the reward (r). Two different definitions of the reward have been considered for the present investigation:

Case 1: purely based on the goal of minimizing the drag coefficient of the manipulated case, hence maximizing the reward defined as

$$r = Cd_0 - Cd, \tag{7}$$

where Cd_0 is the drag coefficient in the unforced condition and Cd is the drag coefficient in the forced condition.

Case 2: accounting not only for the drag reduction but also for the power budget. This was obtained introducing an appropriately defined penalty term.[7,26]

An analysis of the order of magnitude of the drag reduction and the momentum coefficient reveals that the latter is generally negligible with respect to the former. The penalty parameter k_μ was, therefore, defined as follows:

$$k_\mu = \frac{1}{10}\sum_{j=1}^{N_{jets}} \frac{V_k - \bar{V}_j}{V_{\max}}, \tag{8}$$

where V_k and V_{max} are, respectively, equal to 6.5 and 9.5 V, respectively. These values are defined as to ensure that k_μ assumes values within ± 1, which is generally desirable from the convergence of the algorithm; $\bar{V}_j$ is the mean value of the forcing signal for the jth jet over the actuation time. The coefficient 1/10 is a conveniently introduced scaling factor, as to make sure that k_μ and ΔCd have a similar order of magnitude. It is worth explicitly noting that the reward could not be defined according to Eq. (6) since the value of the momentum coefficient of the jets might attain values equal to 0, hence leading to singularities in the reward.

The reward in case 2 is, therefore, defined as

$$r = \Delta Cd + k_\mu. \tag{9}$$

Further precautions were taken to avoid that the agent would learn strategies to maximize the drag reduction alone (ΔCd) or the energy content alone (k_μ). This is not desirable and, as a consequence, the reward was further tuned to promote solutions that would prefer large values of both terms in Eq. (9). In particular, provided that both the rewards components are positive, forcing conditions that lead to differences between the ΔC_d and k_μ that are smaller than a parameter δ (set to 5×10^{-3}) were favored by doubling the reward. Conversely, when the difference between the two rewards was larger than δ, a penalization term was included equal to 0.75. This is summarized in Fig. 3.

IV. RESULTS

In this section, the results obtained from the training of the algorithm are reported and discussed.

The experiments conducted involved different definitions of the state and different definitions of the reward, as summarized in Table III. The goal is to demonstrate that the DRL can achieve satisfactory performance in defining optimal control laws in the experimental environment, with particular attention to understanding the sensitivity of the agent to the state and to evaluating its ability to maximize the reward under different definitions. All the tests are conducted at $Re \sim 10^5$, where Re is defined as

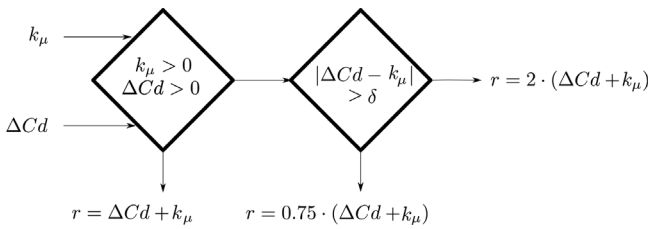


FIG. 3. Algorithm for calculating the reward.

Published under an exclusive license by AIP Publishing

$$Re = \frac{U_\infty \cdot L_{ref}}{\nu}, \tag{10}$$

with $L_{ref} = \frac{H+W}{2} = 0.185$ m and $U_\infty = 9.1$ m/s being the freestream speed.

In the following, each definition of the state (agents A and B) will be investigated against the different definitions of the reward (cases 1 and 2).

A. Training of the agent A

Agent A is characterized by a state defined by the 29 values of the pressure coefficient measured across the model's base.

Figures 4(a) and 4(b) show the evolution of the reward of the DRL algorithm as a function of the episodes for the cases 1 and 2 definitions of the reward, respectively. In particular, the maximum, minimum, and mean reward attained at each episode, i.e., across the 15 timesteps of each episode, are reported in blue, red, and black, respectively.

In both cases, the agent is capable of learning strategies to optimize the reward within about 40 episodes. This corresponds to an experimental time of about 5 h. Beyond that point, the reward attains a plateau, which suggests that the algorithm has reached a sufficient convergence level, and we consider the agent as trained. It is worth explicitly mentioning that an increase in the number of episodes might

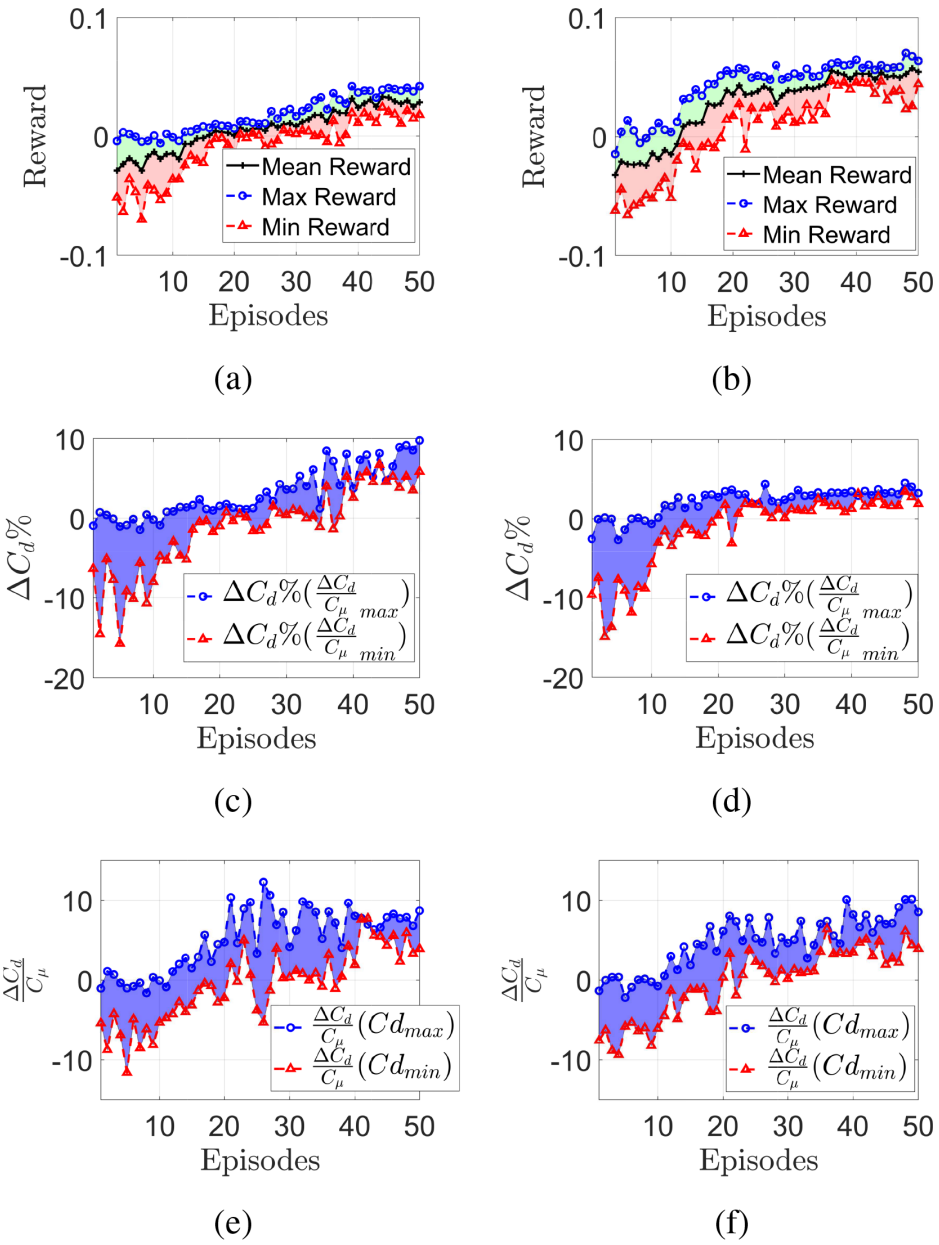


FIG. 4. Evolution of the training of agent A with the reward definitions of case 1 (left) and case 2 (right). (a) and (b) The reward trend, (c) and (d) the percentage variation of the drag coefficient, and (e) and (f) the energy budget following Eq. (6).

Published under an exclusive license by AIP Publishing

lead to further increment of the reward. At the early stages of the training, the algorithm shows a broad exploration of the parametric space, which leads to several forcing conditions that do not yield a positive reward.

The corresponding values of the drag reduction are reported in Figs. 4(c) and 4(d) for cases 1 and 2, respectively. In particular, among all the possible forcing conditions, the evolution of the $\Delta Cd\%$ was assessed in the condition of minimum and maximum value of $\Delta Cd/C_\mu$, indicated in red and blue, respectively. This is done to provide the best possible comparison between the two cases, although in case 1, the $\Delta Cd/C_\mu$ was not accounted for within the reward.

Case 1, which is targeted to the optimization of the drag coefficient alone, leads to values of the reduction as large as $\sim$10% [Fig. 4(c)]. These values are comparable with those obtained in the case of continuous forcing evidenced in previous investigations,[25] but with a significantly larger value of $\Delta Cd/C_\mu$. The continuous jet case was, indeed, characterized by values of drag reduction as large as 11%, but with values of $\Delta Cd/C_\mu < 1$.

Conversely, implementing a pulsed forcing is reflected in the fact that the values of $\Delta Cd/C_\mu$ [Fig. 4(e)] suggest an efficient forcing, with values as large as 7.5 for the maximum drag reduction case.

Case 2, where the power budget is accounted for, attains instead values of drag reduction limited to about 5% [Fig. 4(d)] at the end of the training. However, the corresponding value of the forcing efficiency is consistently larger ($\sim$10).

Figure 5 shows the evolution of the action parameters (amplitude and frequency) as defined in Eqs. (4) and (5), respectively. The symbols in the figures indicate the mean value of the forcing at each episode, while the shadowed area indicates the standard deviation across each episode.

A first immediate comment that can be drawn from the analysis of the forcing parameters is that regardless of the definition of the reward (cases 1 and 2), the agent acts as to switch off the top jet. This result was also confirmed by previous open-loop experiments that showed the detrimental effect of this specific jet on the drag reduction.[25]

All the parameters also show initial larger values of the standard deviation, indicating the exploration of the action space. At the later stages of the training process, these values are progressively reduced, confirming the convergence of the algorithm. The only exception is related to the frequency of the top jet, which, in case 1, features a larger scatter around the mean value. It must be noted though that this jet is progressively switched off during the training, hence the frequency values have little impact on the wake dynamics.

During the training of case 1, it is worth evidencing that the amplitude A_e of the lateral and bottom jet have similar values up until episode 30. Beyond this point, the agent recognizes that the optimal strategy leads to maximizing the amplitude of the lateral jets while keeping values of $A_e \approx -0.5$ for the bottom one. Accounting also for the values of the actuation frequency, case 1 is characterized by a lateral jet with steady forcing and maximum value of the amplitude, while a pulsed jet with a high value of the frequency.

The training of case 2 instead leads to a forcing condition with the lateral and bottom jets characterized by similar values of the

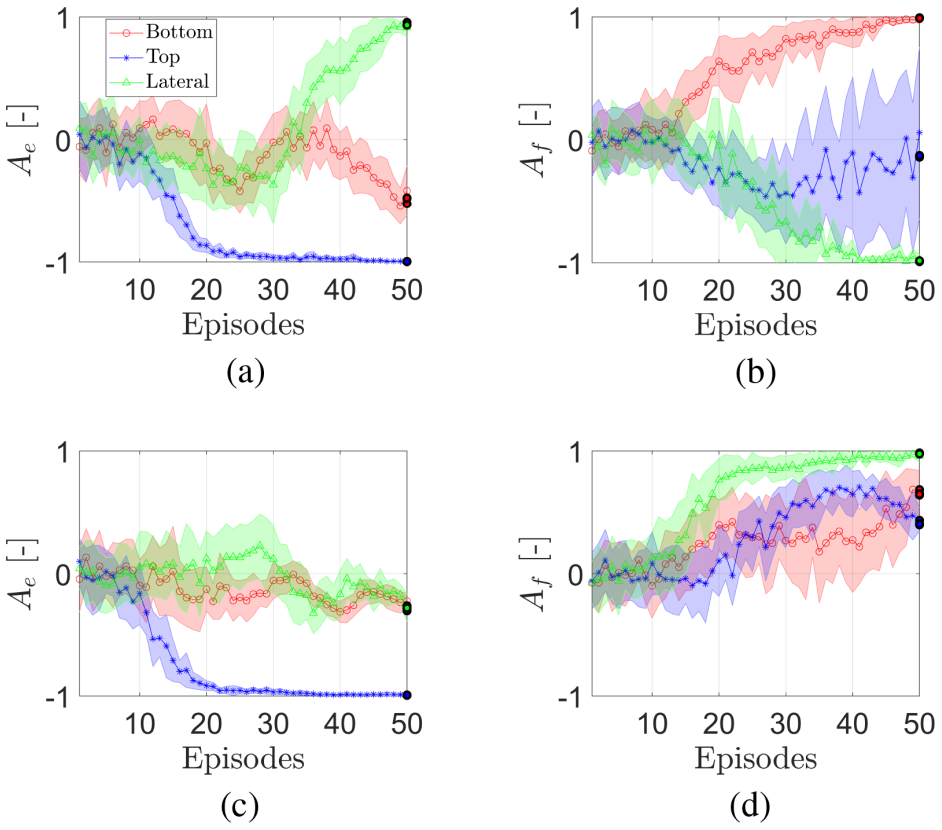


FIG. 5. Mean value of the action parameters for agent A: amplitude on the left and frequencies on the right. The first row (a) and (b) corresponds to case 1, while the second row (c) and (d) corresponds to case 2. The bold symbols at the final step indicate the values of the forcing parameters used during the evaluation.

Published under an exclusive license by AIP Publishing

amplitude and frequency. The frequency of the bottom and lateral jets tends toward high values ($A_f \rightarrow 1$). It appears from Figs. 4(c) and 4(d) that for an efficient control of the wake, the control law converges toward moderate values of A_e large actuation frequencies.

B. Training of the agent B

As already mentioned in Sec. III, the effect of a broader parametric space has been investigated by expanding the state definition to account also for the fluctuating pressure signals measured across the model's base. This has been embodied in the framework by including the values of the Cp_{RMS} values measured across the base, thus effectively extending the number of inputs that define the state to 43.

Similarly to the agent A, the agent B was trained with two different reward definitions, as reported in Eqs. (7) and (9).

Figures 6(a) and 6(b) show the training of the agent B for the two definitions of the reward, cases 1 and 2, respectively. The effect of the broader parametric space is expected to be twofold: first, a quicker convergence toward the optimal solution should be attained; second, the solutions should be less prone to suffer from local minima effects.

The first condition is, indeed, verified: the agent attains a constant value of the reward within 30 episodes in case 1, while even within a smaller number of episodes in case 2.

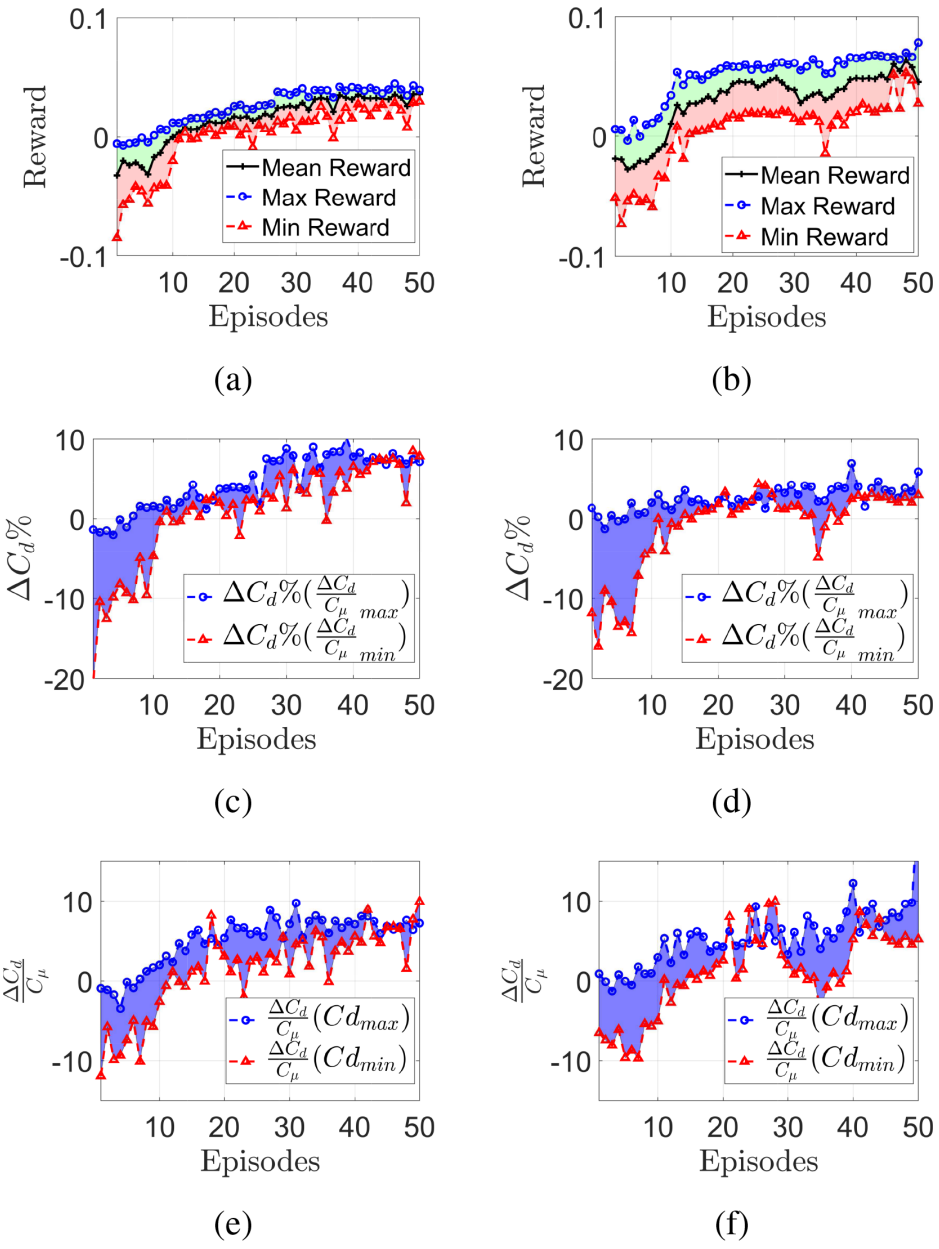


FIG. 6. Evolution of the training of agent B with the reward definitions of case 1 (left) and case 2 (right). (a) and (b) The reward trend, (c) and (d) the percentage variation of the drag coefficient, and (e) and (f) the energy budget following Eq. (6).

Published under an exclusive license by AIP Publishing

The corresponding values of the drag reduction are reported in Figs. 6(c) and 6(d). The maximum values of drag reduction that the agent is capable of attaining, in this case, are similar to or slightly greater than the case of agent A. While the number of probes has primarily an effect on the convergence time of the system, little influence can be identified on the value of the reward. Hence, case 1 is characterized by a maximum drag reduction of 10%, while case 2 attains a drag reduction of 6%.

This result would suggest the independence of the outcome from the size of the state. It should be noted, however, that the increase in-state space is about ∼50% (from 29 values to 43 values). Moreover, the resulting forcing conditions learned by the two agents are different, hence representing two possible solutions that are characterized by a similar value of the reward. This suggests that for each case, the agent has brought itself into a condition of a local maximum.

Figure 7 shows the evolution of the action parameters as a function of the evaluation steps. Consistently with the results of agent A, also the agent trained with the full state leads to a solution where the top jet is shut down at the end of the training phase. Furthermore, it is interesting to notice that the values of the forcing that stem from the training are not too different, at least in terms of the amplitude A_e between the two agents. In particular, for case 1, the lateral jet is pushed toward solutions with large values of the amplitude, whereas the bottom jet is characterized by values of the amplitude that are close to the mean value. It is also worth mentioning that initially the bottom and lateral jets follow similar values of the actuation, until episode 10. Beyond this point, the agent tends to maximize the actuation voltage for the lateral jet, while keeping a value of $A_e \approx -0.25$ for the bottom one. Also in the case of agent B, the bottom and lateral jet toward values of A_f are corresponding to maximum frequency and steady forcing, respectively.

The main difference between the solutions of the two agents in terms of the amplitude of the forcing can be detected in case 2. In particular, agent B would lead to solutions where the control is mainly operated by the lateral jets [Fig. 7(c)]. On the other hand, agent A shows values of the amplitude that are similar for both the bottom and the lateral jets [Fig. 5(c)]. Since the reward of these solutions is similar, this suggests that they are both possible solutions, representing local minima for the system. It is also worth noticing that during the training, the values of A_e for the lateral and bottom jets are typically opposite in behavior: an increase in the bottom jet actuation corresponds to a decrease in the lateral one and vice versa. This behavior is likely to be related to the choice of the reward that also keeps into account the energy spent in the actuation.

C. Evaluation of the agents

The trained agents A and B were then evaluated, both for cases 1 and 2. The evaluation differs from the training in that the agent is prevented from continuing the exploration of the parametric space and the update of the weights of the neural network.

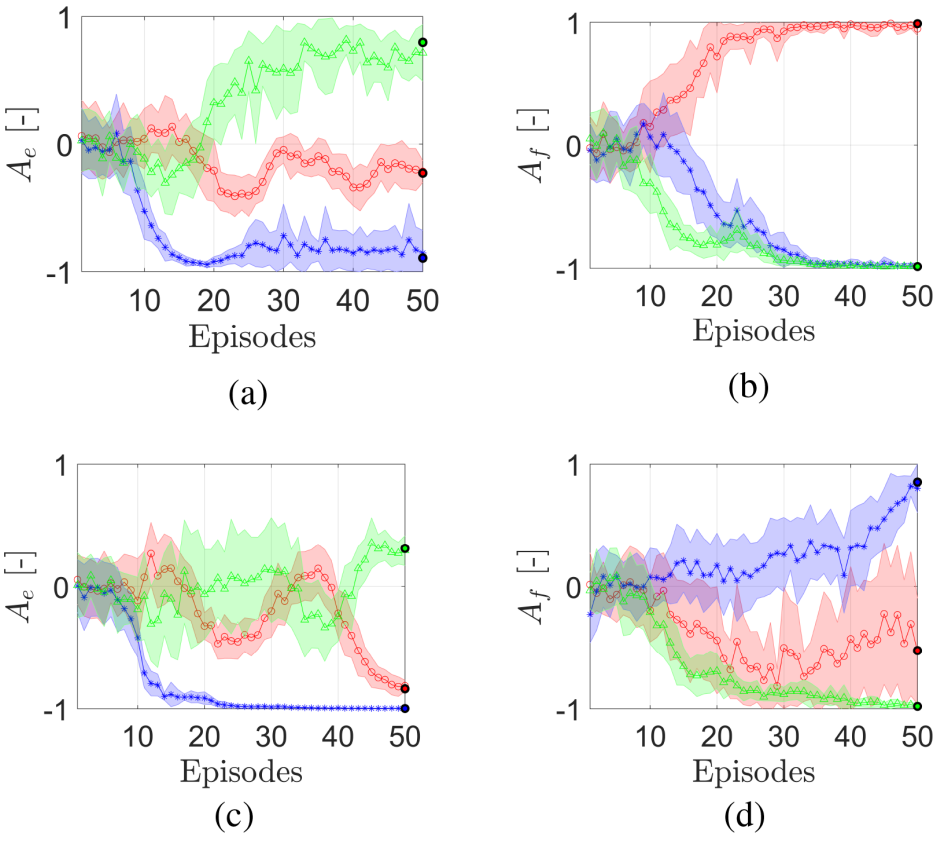


FIG. 7. Mean value of the action parameters for agent B: amplitude on the left and frequencies on the right. The first row (a) and (b) corresponds to case 1, while the second row (c) and (d) corresponds to case 2. The bold symbols at the final step indicate the values of the forcing parameters used during the evaluation.

Published under an exclusive license by AIP Publishing

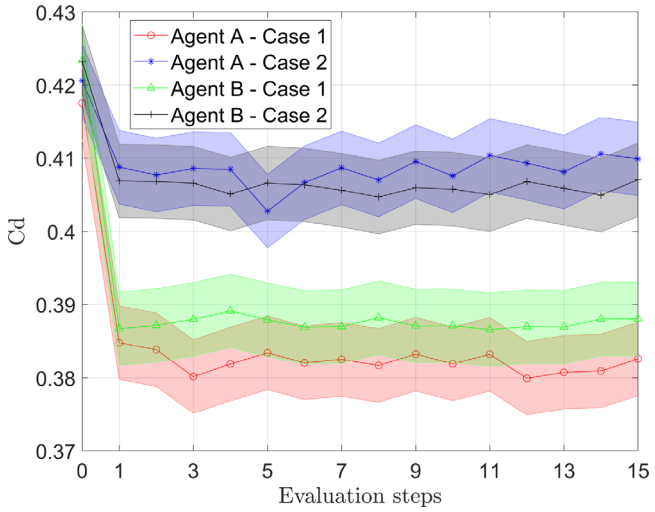


FIG. 8. Evaluation of the drag coefficient when controlled by the trained agents A and B. The shadowed area indicates the uncertainty bars of the measured values (±1%).

Figure 8 shows the drag coefficient measured during the evaluation phase. The first evaluation step corresponds to the unforced condition. After that step, the agent starts controlling the wake. The summary of the forcing conditions expressed in terms of the momentum coefficient and normalized frequency is reported in Tables IV and V for the agent A and B, respectively.

The results show that regardless of the agent and the reward definition (case 1 or 2), the trained agents can effectively control the wake within one evaluation step. In particular, the agents trained with the reward defined to minimize drag, case 1, achieve values of drag reduction of approximately 9%; conversely, the agent trained to minimize the reward defined in Eq. (9), case 2, attains values of the drag reduction of about 4.5%.

The differences between the two agents are, as evidenced in the figure, within the uncertainty of the measurements.

Further insights into the results can be also inferred from Figs. 9(a) and 9(b), where the percentage drag reduction is reported along with the wake receptivity, $\Delta Cd/C_\mu$, for agents A and B, respectively. The results show that consistently with the definition of the reward, the agents prioritize the drag reduction (in red) or the $\Delta Cd/C_\mu$. Interestingly, even conditions that maximize the drag reduction alone, result to be favorable from the efficiency perspective. Nevertheless, a reward specifically tailored to improve the efficiency

TABLE IV. Summary of the forcing parameters expressed in terms of the momentum coefficient (C_μ) and normalized frequency (fW/U_∞) for the top, bottom, and lateral jets obtained in the evaluation phase with agent A.

	Top		Bottom		Lateral	
	$C_\mu \times 10^{-3}$	fW/U_∞	$C_\mu \times 10^{-3}$	fW/U_∞	$C_\mu \times 10^{-3}$	fW/U_∞
Case 1	0.0	0.24	0.55	0.55	1.65	0.004
Case 2	0.0	0.39	1.16	0.45	0.18	0.54

TABLE V. Summary of the forcing parameters expressed in terms of the momentum coefficient (C_μ) and normalized frequency (fW/U_∞) for the top, bottom, and lateral jets obtained in the evaluation phase with agent B.

	Top		Bottom		Lateral	
	$C_\mu \times 10^{-3}$	fW/U_∞	$C_\mu \times 10^{-3}$	fW/U_∞	$C_\mu \times 10^{-3}$	fW/U_∞
Case 1	∼0	0.004	1.29	0.55	1.42	0.004
Case 2	0.0	0.51	0.03	0.13	0.71	0.006

leads to increments of $\Delta Cd/C_\mu$ of about 45% with respect to case 1. The larger oscillations that are evidenced in the $\Delta Cd/C_\mu$ ought to be expected, given the strong sensitivity to small variations of the momentum coefficient. The summary of the resulting values of ΔCd and the $\Delta Cd/C_\mu$ obtained during the evaluation of the two agents A and B for the two definitions of the reward is reported in Table VI.

As it is summarized in Tables IV and V, the values of the momentum coefficient are of the order of $\mathcal{O}(10^{-3})$, hence small variations are immediately reflected into the value of the $\Delta Cd/C_\mu$.

D. Effect of the forcing on the wake

The effects of the forcing on the structure of the near wake are investigated in terms of the pressure distribution across the model's base.

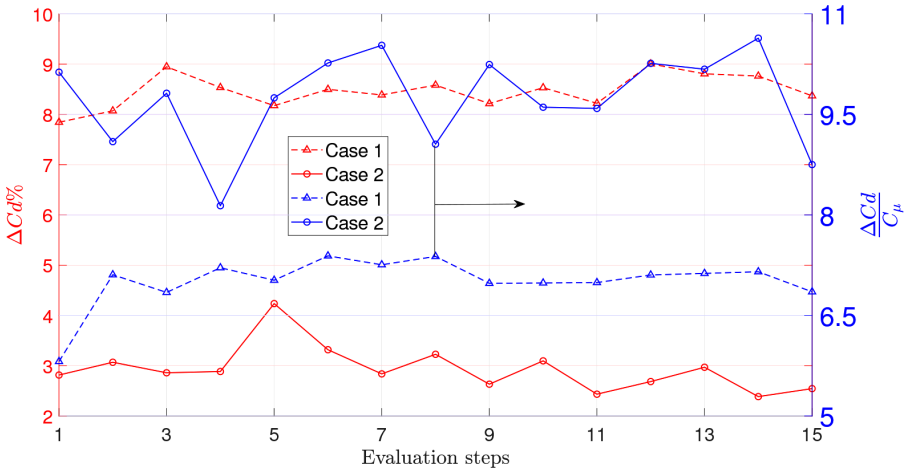
Figure 10 shows the static pressure coefficient distribution comparing the unforced and forced conditions for each agent and case analyzed. The Cp distribution in the natural case has been widely investigated in the literature.[25,27,38] The signature of a large recirculating bubble that originates from the top edge of the vehicle and extends for more than half of the model's height is reflected into a region of low pressure above $Z/W > 0.4$, with a downward directed pressure gradient. The spatially averaged pressure coefficient calculated across the base is equal to $\overline{Cp}_{base} = -0.12$, in good agreement with previous measurements in the literature, for similar values of the aspect ratio of the model's base and of the Reynolds number.

The controlled cases are instead characterized by a significantly different structure of the near wake. A first qualitative observation suggests that despite the solutions in terms of reward being similar regardless of the agent and the reward definition, the near wake structure can be quite different.

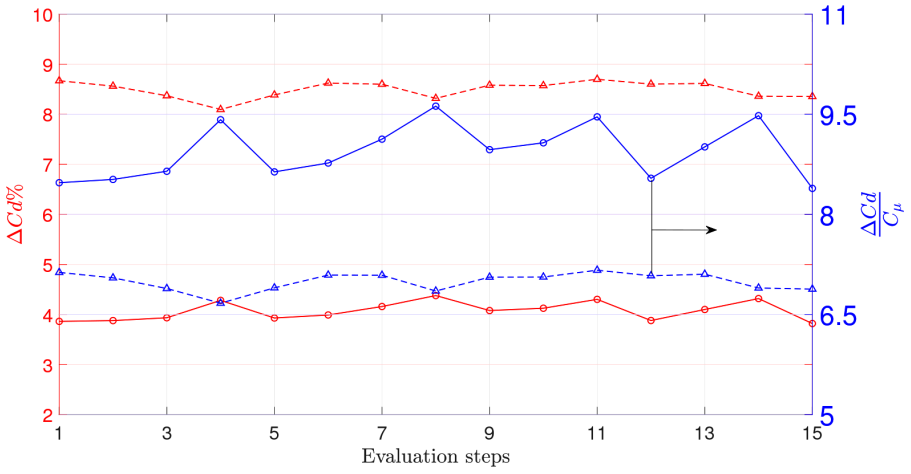
Agent A, in case 1, attains a condition of the wake characterized by greater values of the pressure coefficient in the middle of the base, at $Z/W = 0.5$. This suggests that the large recirculation bubble that originates from the top edge of the model is now reorganized into smaller structures from the top, lateral, and bottom edges. The corresponding value of the mean pressure coefficient is $\overline{Cp}_{base} = -0.07$, with a significant pressure recovery.

The relatively small differences between the forcing parameters in cases 1 and 2 for agent A are also reflected in a similar structure of the near wake, at least from the topology point of view. The reduced values of A_e for the lateral jet yield a less expensive, in terms of power budget, forcing but characterized by lower drag reduction. The mean pressure coefficient calculated across the rear base is $\overline{Cp}_{base} = -0.10$, with only limited pressure recovery.

Published under an exclusive license by AIP Publishing



(a)



(b)

FIG. 9. Evaluation of the agents' performance in terms of percentage drag reduction $\Delta Cd\%$ (in red) and wake receptivity $\frac{\Delta Cd}{C_\mu}$ (in blue). (a) Agent A and (b) agent B.

The resulting structure of the wake controlled by agent B shows a behavior that does not resemble the results obtained with agent A, despite the resulting values of the drag coefficient and the efficiency being quite similar. In case 1, the effect of the bottom jet is such that the wake features a reversed pressure gradient compared to the natural case. Following the interpretation given for agent A, this suggests that the shear layer originating from the top edge and rolling up into a recirculating bubble is now significantly smaller in size. This is a behavior that was already observed in the case of maximum drag reduction when the wake was forced with continuous jets.[25] Similar values to the agent A, case 1 of the spatially averaged pressure coefficient, are also attained in this case, $\overline{Cp}_{base} = -0.075$, thus justifying the similar values of drag reduction.

TABLE VI. Summary of the $\Delta Cd\%$ and the $\Delta Cd/C_\mu$ obtained during the evaluation (or during the training in the parentheses) of agents A and B for cases 1 and 2.

	Agent A		Agent B	
	$\Delta Cd\%$	$\Delta Cd/C_\mu$	$\Delta Cd\%$	$\Delta Cd/C_\mu$
Case 1	8.46 (9.15)	7.01 (8.26)	8.49 (8.51)	6.99 (9.92)
Case 2	3.03 (4.45)	9.73 (10.16)	4.07 (5.61)	8.94 (9.65)

The last case corresponds to agent B, case 2. Consistently with the less intense forcing of the wake, the pressure coefficient distribution resembles closely the natural condition, with a more extended high pressure region that reaches $Z/W = 0.5$. Despite the different structure of the wake, also in this case, the value of $\overline{Cp}_{base} = -0.095$, which is not too different from the result obtained by agent A, case 2.

V. CONCLUSIONS

Deep Reinforcement Learning was implemented in an experimental framework to learn optimal strategies to minimize the drag of a model reproducing a road vehicle.

Published under an exclusive license by AIP Publishing

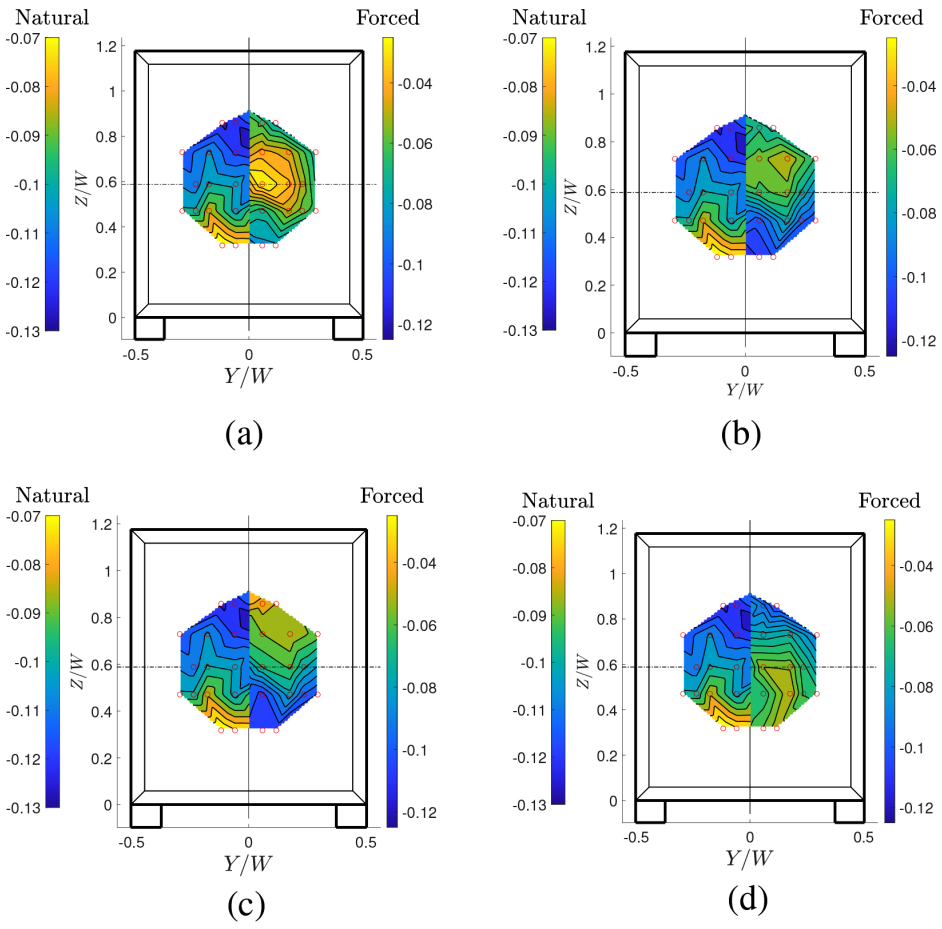


FIG. 10. Pressure coefficient distribution *Cp* across the rear of the model in the natural (left) and forced (right) conditions for the four investigated cases. Agent A, case 1 (a); agent A, case 2 (b); agent B, case 1 (c); agent B, case 2 (d).

The control of the wake was operated using pulsed jets located along the edges of the base of the model.

The analysis was focused on two main aspects: (i) the number of elements that constitute the environment, i.e., the representation of the state; (ii) the definition of the reward.

The results showed that an increase in the number of probes, from 29 static pressure sensors to 43 static and fluctuating pressure sensors, has a direct effect on the time required by the agent to learn the optimal strategy. On the other hand, no significant effect has been detected on the value of the reward. Indeed, both agents A and B are capable of attaining values of the drag reduction that are of the order of 10%, when the reward is purely aimed at minimizing drag (case 1), while about 4.5% when the reward takes into account a figure of the efficiency of the forcing, defined through the wake receptivity $\Delta Cd/C_\mu$ (case 2).

The resulting forcing condition that minimizes the drag for the different agents and rewards defined are listed in Tables IV and V, respectively. The consistent result across all the investigated cases is the detrimental effect of the top jet on the drag reduction. The actuation frequencies, on the other hand, seem to suggest that the lateral jet is more efficient and effective when operated in nearly steady conditions. This is true for all the cases except agent A, case 2. The bottom jet is generally operated at generally large values of the frequency, with the specific case of agent 2, case 2, where the non-dimensional frequency of 0.13 matches exactly the shedding frequency for the problem. These aspects will be further investigated in future work with flow field measurements.

It is interesting to notice that, comparing the results obtained in the present investigations to the ones obtained when forcing with a continuous jet system,[25] even the solutions yielding the maximum drag reduction are sound from the efficiency point of view. In particular, cases leading to the maximum drag reduction can obtain values of the efficiency as large as 4, whereas the value increases to approximately 9 when the efficiency is kept into account.

The analysis of the pressure coefficient measured across the base of the model reveals that the two agents converge to solutions that represent local minima for the system, although characterized by similar values of the reward.

The results obtained in the present investigation prompt further analysis of the possible implementation of the technique for the efficient drag reduction of road vehicles. In particular, the actuation of the air jets could be implemented exploiting the exhaust gases of the engine through a thermodynamic cycle.[39]

Published under an exclusive license by AIP Publishing

ACKNOWLEDGMENTS

E.A. acknowledges the CEUR foundation for supporting his research. The authors acknowledge Alexandre Kuhnle for the fruitful discussions at the early stages of the implementation.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Enrico Amico: Data curation (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (equal). **Gioacchino Cafiero:** Conceptualization (equal); Methodology (equal); Software (equal); Writing – review & editing (equal). **Gaetano Iuso:** Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: EFFICIENCY FOR AN AIR-BREATHING SYSTEM

As already mentioned in Sec. III A, the definition of an appropriate figure for the efficiency is not an easy task. In this paper, the wake receptivity has been considered ($\Delta Cd/C_\mu$), since the feeding system is based on a pressurized tank. Nonetheless, among the others, a possible actuation might be based on feeding the system with a dedicated air intake connected with the freestream condition. This would resemble the case of an air breathing engine, which might potentially lead to greater values of the efficiency, or at least, reduce the burden on the compressor. In this specific case, it is possible to introduce the ratio of the power saved $\Delta D U_\infty$ over the power variation between the jet exit section and the freestream conditions as

$$\eta_\infty = \frac{2\Delta D U_\infty}{\sum_{j=1}^{N_{jets}} \dot{m}_j (U_j^2 - U_\infty^2)}, \tag{A1}$$

where $\dot{m}_j$ represents the mass flow rate of the jth jet. It must be noted though that this definition of the efficiency would be dependent on the appropriate selection of the reference speed.

REFERENCES

[1]P. Gilliéron and A. Kourta, "Aerodynamic drag reduction by vertical splitter plates," Exp. Fluids **48**, 1–16 (2010).
[2]B. Schoettle, M. Sivak, and M. Tunnell, "A survey of fuel economy and fuel usage by heavy-duty truck fleets," American Transportation Research Institute Report No. SWT-2016-12 (2016).
[3]A. Capone and G. Romano, "Investigation on the effect of horizontal and vertical deflectors on the near-wake of a square-back car model," J. Wind Eng. Ind. Aerodyn. **185**, 57–64 (2019).
[4]P. Gehlert, Z. Cherfane, G. Cafiero, and J. C. Vassilicos, "Effect of multiscale endplates on wing-tip vortex," AIAA J. **59**, 1614 (2021).
[5]G. Minelli, T. Dong, B. R. Noack, and S. Krajnović, "Upstream actuation for bluff-body wake control driven by a genetically inspired optimization," J. Fluid Mech. **893**, A1 (2020).
[6]R. Castellanos, G. Y. Cornejo Maceda, I. de la Fuente, B. R. Noack, A. Ianiro, and S. Discetti, "Machine-learning flow control with few sensor feedback and measurement noise," Phys. Fluids **34**, 047118 (2022).
[7]J. Rabault, M. Kuchta, A. Jensen, U. Reglade, and N. Cerardi, "Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control," J. Fluid Mech. **865**, 281 (2019).
[8]G. Pujals, S. Depardon, and C. Cossu, "Drag reduction of a 3D bluff body using coherent streamwise streaks," Exp. Fluids **49**, 1085–1094 (2010).
[9]J.-F. Beaudoin and J.-L. Aider, "Drag and lift reduction of a 3D bluff body using flaps," Exp. Fluids **44**, 491–501 (2008).
[10]D. Kim, H. Lee, W. Yi, and H. Choi, "A bio-inspired device for drag reduction on a three-dimensional model vehicle," Bioinspiration Biomimetics **11**, 026004 (2016).
[11]G. Iuso, "Base pressure control by passive methods," in 18th Fluid Dynamics and Plasmadynamics and Lasers Conference, 1985.
[12]E. A. Gillies, "Low-dimensional control of the circular cylinder wake," J. Fluid Mech. **371**, 157–178 (1998).
[13]O. Stalnov, I. Fono, and A. Seifert, "Closed-loop bluff-body wake stabilization via fluidic excitation," Theor. Comput. Fluid Dyn. **25**(1), 209–219 (2011).
[14]L. D. Longa, A. S. Morgans, and J. A. Dahan, "Reducing the pressure drag of a d-shaped bluff body using linear feedback control," Theor. Comput. Fluid Dyn. **31**(5), 567–577 (2017).
[15]D. Lasagna, M. Orazi, and G. Iuso, "Multi-time delay, multi-point linear stochastic estimation of a cavity shear layer velocity from wall-pressure measurements," Phys. Fluids **25**, 017101 (2013).
[16]D. Lasagna, L. Fronges, M. Orazi, and G. Iuso, "Nonlinear multi-time-delay stochastic estimation: Application to cavity flow and turbulent channel flow," AIAA J. **53**, 2920–2935 (2015).
[17]E. Amico, D. D. Bari, G. Cafiero, and G. Iuso, "Genetic algorithm-based control of the wake of a bluff body," J. Phys.: Conf. Ser. **2293**, 012016 (2022).
[18]Y. Zhou, Y. Zhou, D. Fan, B. Zhang, R. Li, R. Li, B. R. Noack, B. R. Noack, and B. R. Noack, "Artificial intelligence control of a turbulent jet," J. Fluid Mech. **897**, 27 (2020).
[19]R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, 2015).
[20]R. Paris, S. Beneddine, and J. Dandois, "Robust flow control and optimal sensor placement using deep reinforcement learning," J. Fluid Mech. **913**, A25 (2021).
[21]F. Ren, J. Rabault, and H. Tang, "Applying deep reinforcement learning to active flow control in weakly turbulent conditions," Phys. Fluids **33**, 037121 (2021).
[22]P. Garnier, J. Viquerat, J. Rabault, A. Larcher, A. Kuhnle, and E. Hachem, "A review on deep reinforcement learning for fluid mechanics," Comput. Fluids **225**, 104973 (2021).
[23]F. Pino, L. Schena, J. Rabault, A. Kuhnle, and M. A. Mendez, "Comparative analysis of machine learning methods for active flow control," arXiv:2202.11664 (2022).
[24]J. Viquerat, P. Meliga, and E. Hachem, "A review on deep reinforcement learning for fluid mechanics: An update," arXiv:2107.12206 (2021).
[25]J. J. Cerutti, C. Sardu, G. Cafiero, and G. Iuso, "Active flow control on a square-back road vehicle," Fluids **5**, 55 (2020).
[26]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," Proc. Natl. Acad. Sci. U. S. A. **117**, 26091–26098 (2020).
[27]J. J. Cerutti, G. Cafiero, and G. Iuso, "Aerodynamic drag reduction by means of platooning configurations of light commercial vehicles: A flow field analysis," Int. J. Heat Fluid Flow **90**, 108823 (2021).
[28]D. Barros, J. Borée, B. R. Noack, A. Spohn, and T. Ruiz, "Bluff body drag manipulation using pulsed jets and coanda effect," J. Fluid Mech. **805**, 422 (2016).
[29]T. Castelain, M. Michard, M. Szmigiel, D. Chacaton, and D. Juvé, "Identification of flow classes in the wake of a simplified truck model depending on the underbody velocity," J. Wind Eng. Ind. Aerodyn. **175**, 352–363 (2018).

Published under an exclusive license by AIP Publishing

[30]C. Sardu, D. Lasagna, and G. Iuso, "Noise filtering for wall pressure fluctuations in measurements around a cylinder with laminar and turbulent flow separation," J. Fluids Eng. **138**, 061101 (2016).
[31]A. Kuhnle, M. Schaarschmidt, and K. Fricke, "Tensorforce: A tensorflow library for applied reinforcement learning," https://tensorforce.readthedocs.io/en/latest/ (2017).
[32]J. Viquerat, J. Rabault, A. Kuhnle, H. Ghraieb, A. Larcher, and E. Hachem, "Direct shape optimization through deep reinforcement learning," J. Comput. Phys. **428**, 110080 (2021).
[33]E. Hachem, H. Ghraieb, J. Viquerat, A. Larcher, and P. Meliga, "Deep reinforcement learning for the control of conjugate heat transfer," J. Comput. Phys. **436**, 110317 (2021).
[34]H. Ghraieb, J. Viquerat, A. Larcher, P. Meliga, and E. Hachem, "Single-step deep reinforcement learning for two- and three-dimensional optimal shape design," AIP Adv. **12**, 085108 (2022).
[35]H. Dong, Z. Ding, and S. Zhang, *Deep Reinforcement Learning: Fundamentals Research and Applications* (Springer, Singapore, 2017).
[36]H. Choi, W.-P. Jeon, and J. Kim, "Control of flow over a bluff body," Annu. Rev. Fluid Mech. **40**, 113–139 (2008).
[37]R. Englar, "Advanced aerodynamic devices to improve the performance, economics, handling, and safety of heavy vehicles," SAE Technical Paper No. 2001-01-2072 (2001).
[38]M. Grandemange, M. Gohlke, and O. Cadot, "Turbulent wake past a three-dimensional blunt body. Part 1. Global modes and bi-stability," J. Fluid Mech. **722**, 51 (2013).
[39]V. Tesar, A. Konig, J. Macek, and P. Baumruk, "New ways of fluid flow control in automobiles: Experience with exhaust gas aftertreatment control," Seoul 2000 FISITA World Automotive Congress, 12-15 June 2000, Seoul, Korea (2000).

Published under an exclusive license by AIP Publishing