13th International Symposium on Turbulence and Shear Flow Phenomena (TSFP13)
Montreal, Canada, June 25–28, 2024

INFLUENCE OF THE STATE OBSERVATION ON DEEP-REINFORCEMENT-LEARNING DRAG-REDUCTION POLICIES IN WALL-BOUNDED FLOWS

Ricardo Vinuesa
FLOW, KTH Engineering Mechanics,
SE-100 44 Stockholm, Sweden
rvinuesa@mech.kth.se

Jean Rabault
Independent researcher,
Oslo, Norway
jean.rblt@gmail.com

Hossein Azizpour
School Elect. Eng. and Comp. Sci., KTH
SE-100 44 Stockholm, Sweden
azizpour@kth.se

Luca Guastoni
FLOW, KTH Engineering Mechanics,
SE-100 44 Stockholm, Sweden
guastoni@mech.kth.se

ABSTRACT

The control of turbulent fluid flows represents a problem in several engineering applications. The chaotic, high-dimensional, non-linear nature of turbulence makes it difficult to design control strategies with traditional frameworks. Recently, deep reinforcement learning (DRL) has been shown to be an effective approach to obtain control policies that reduce the friction drag in different wall-bounded turbulent flows. In this work we compare different drag-reduction strategies that compute their actuation based on the fluctuations at a given wall-normal location in turbulent open channel flow. In particular, we consider the policies learnt using deep reinforcement learning (DRL) based on sensing planes at the following inner-scaled locations: $y_s^+ = (10, 15, 30, 50)$, and compare them with opposition control (Choi *et al.*, 1994). By using the deep deterministic policy gradient (DDPG) algorithm, we are able to discover control strategies that outperform existing control methods at two inner-scaled locations closer to the wall. When using information farther from the wall, opposition control increases the drag. By contrast, DRL is able to learn drag-reducing policies in a minimal channel flow using observations at $y_s^+ = 30$ and 50.

INTRODUCTION

Deep reinforcement learning (DRL) has recently emerged as a promising technique to learn novel, adaptive flow-control strategies in an automated way (Vignon *et al.*, 2023). This is particularly relevant for discovering drag-reduction policies, as they may be applied in different industries, such as aerospace, automotive, and marine engineering. Recent works have applied the reinforcement-learning framework in numerical simulations of different flow cases: the flow around a cylinder has been studied both in 2D (Rabault *et al.*, 2019) and 3D (Suárez *et al.*, 2023), obtaining different policies depending on the Reynolds number investigated (Varela *et al.*, 2022). DRL has enabled the discovery of drag-reducing policies in three-dimensional wall-bounded cases such as Couette (Zeng *et al.*, 2022) and Poiseuille flows (Guastoni *et al.*, 2023; Sonoda *et al.*, 2023). DRL has also been used to reduce the size of a recirculation bubble in a turbulent boundary layer (Font *et al.*, 2024). Other notable applications in fluid mechanics of this technique are the maximization of the efficiency of agents swimming in a turbulent flow (Biferale *et al.*, 2019; Verma *et al.*, 2018), mesh optimization (Lorsung & Barati Farimani, 2023) and turbulence modelling (Novati *et al.*, 2021; Kurz *et al.*, 2023).

In reinforcement learning (RL), a controlling agent learns through trial and error how to control the environment, which is characterized at a given time t by a state variable $\mathbf{s}_t \in S$, where S indicate the set of all possible states of the environment. The agent picks an action $a_t \in A$ and it receives a reward r_t based on the effect of the action just performed. Here A represents the set of possible actions in a given state. In this work, we compare different drag reduction strategies in turbulent channel flow, building on the reinforcement-learning formulation of the problem from our previous study (Guastoni *et al.*, 2023). In particular, we analyze the relationship between the state observation and the learned policy in the context of drag reduction. We compare the results with an established non-data-driven control techniques, namely opposition control (Choi *et al.*, 1994).

The state observation is the only information about the current fluid system state, that is provided to the learning agent to develop a control strategy. The challenge related to this control task is two-fold: first, the agent needs to learn how the action performed affects the reward and the observed state, then it needs to devise a successful strategy to maximize the collected reward. These two tasks are performed at the same time during the learning phase. We consider the observation of the velocity at different wall-normal locations in order to identify which flow features enable the learning of the best performing policies.

METHODOLOGY

As in our previous study (Guastoni *et al.*, 2023) on the application of reinforcement learning for drag reduction, we study a 3D turbulent open channel flow. In order to apply DRL to a fluid-dynamics problem, we need to define the observed state, the actions that can be performed by the agent and the reward function $r = r(s_{t+1}, s_t, a_t) : S^2 \times A \to \mathbb{R}$ that quantifies

the effect of the actions performed by the agent. We provide the streamwise and wall-normal components of the velocity at a given wall-normal location as state observation. The control is implemented as a wall-normal velocity distribution at the wall and the reward is defined as the wall-shear stress variation with respect to the uncontrolled case.

The control problem is formulated as a multi-agent reinforcement learning (MARL) problem. In this setting, a grid of $N_{\mathrm{CTRLx}} \times N_{\mathrm{CTRLz}}$ control agents determines the control actions at the wall. The number of agents depends on the flow simulation. Each agent receives the velocity fluctuations at a given (x,z) position and it computes the action value right below it. Since the flow simulation is incompressible, the average of the control signal is removed to obtain zero-net-mass-flux actions at each timestep.

In this study, we aim to analyze the effect of state observations on the DRL learning process and policy. To this end, we perform several separate learning runs during which we sample the streamwise and wall-normal velocity fluctuations sensing planes located at progressively larger inner-scaled wall-normal locations, namely $y_s^+ = (10, 15, 30, 50)$. Since the control is imposed at the wall, the closer is the sensing plane to the wall, the faster the observed state will be affected. Furthermore, the flow modifications caused by the control will be less evident and localized as we move farther away from the wall. This depends on the combined effect of diffusion and advection in the flow. Overall, higher wall-normal distances for the sensing plane will result in a harder learning task for the agent, as the reward signal becomes noisier as y^+ increases.

We perform the learning process using deep deterministic policy gradient (DDPG) algorithm (Lillicrap *et al.*, 2015) in a minimal open channel flow (Jiménez & Moin, 1991). The learned policy is evaluated in both in a minimal channel (with domain size $\Omega = L_x \times L_y \times L_z = 2.67h \times h \times 0.8h$ and h is the open-channel height) and in a larger channel ($\Omega = L_x \times L_y \times L_z = 2\pi h \times h \times \pi h$), both with uncontrolled friction Reynolds number $Re_\tau = 180$. In the minimal channel flow $N_{\mathrm{CTRLx}} \times N_{\mathrm{CTRLz}} = 16 \times 16$, whereas in the larger channel $N_{\mathrm{CTRLx}} \times N_{\mathrm{CTRLz}} = 64 \times 64$. The results at each wall-normal location in both the minimal channel and larger channel are compared with an opposition control policy that receives as input the velocity field at the corresponding sensing plane y_s^+.

RESULTS

As described more in detail in Guastoni *et al.* (2023), the policy learnt using the state observations at $y_s^+ = 15$ is consistently providing an average drag reduction that is higher than the one obtained with opposition control. After the initial transient ($t^+ > 500$) the DDPG policy provides 43% drag reduction, while opposition control is limited to 26%. Furthermore, having a local and translational invariant policy allows us to apply it to a different domain size with no modifications. This way, we are able to obtain 30% drag reduction in the larger channel, whereas opposition control is limited to 20%. Here we apply the same approach to study the DRL policies obtained by observing the state at different wall-normal locations. All the policies are tested on 6 different initial conditions, both in the minimal and larger channel. The drag reduction comparison of the two control strategies is summarized in figure 1.

First we analyze the performance of the control strategies in the minimal channel. Closer to the wall (at $y_s^+ = 10$ and $y_s^+ = 15$), both control techniques are able to effectively reduce drag. At $y_s^+ = 30$, opposition control barely yields drag reduction and at $y_s^+ = 50$ it generates a substantial drag increase. This is due to the fact that the opposition-control strategy is based on the suppression of the near-wall turbulent structures that are more difficult to observe farther from the wall. By contrast, the same DRL algorithm used in our previous study can be used to learn effective drag-reduction policies in the minimal channel even when the state of the flow is reconstructed using observations at $y_s^+ = 30$ and $y_s^+ = 50$. It is important to note how the performance can differ significantly in the minimal and regular channels: for instance, the DRL policy with input at $y_s^+ = 30$ provides almost 20% drag reduction in the smaller domain, but it is much less effective in the larger one. When moving to $y_s^+ = 50$, the considered policy yields a drag increase in the larger channel. This highlights the importance of testing the DRL policy in both domains, to test how it generalizes.

At all wall-normal positions of the sensing plane, the DRL allows the discovery of drag-reducing policies. With the use of a MARL approach, the observation space of each individual agent is $A \subseteq \mathbb{R}^2$. This allows a two-dimensional representation of the learnt policy. Since we are considering the streamwise and wall-normal fluctuations as inputs, this representation is similar to the one obtained through quadrant analysis (Wallace *et al.*, 1972). We compare the best-performing policies obtained with different observation sampling planes y_s^+ in figure 2. $y_s^+ = 10$ and $y_s^+ = 15$ are close and the RL policies are also qualitatively similar. Differently from opposition control, the action mainly depends on the sign of the streamwise velocity fluctuations, while the intensity and sign of the wall-normal velocity fluctuations do not affect the actions. The control sign changes abruptly as the u-fluctuations sign changes. The policy learnt with $y_s^+ = 30$ exhibits a similar dependency on the velocity fluctuations, but the variation with the streamwise component intensity is smoother. The policy with input at $y_s^+ = 50$ conceptually different, focusing on the Q4 ($+u'$, $-v'$) quadrant. It should be noted that the RL optimizes the policy in a *black-box* fashion. This prevent us from verifying whether the policy learning can be potentially improved further and whether the improvement is hindered by the observability of the action effects from the state or in modelling the reward signal (*i.e.* deciding which action to perform based on the state observation). It is important to highlight that a naïve application of the policy learnt at $y_s^+ = 15$ using the observations father away from the wall will result in a drag increase.

After analyzing the actions that are performed as a function of the observed state, we consider another approach to compare the classical and learnt policies, by monitoring the distribution of the velocity fluctuations when the control is applied. Independently from the sampling plane y_s^+ used by the policy, we consider the same wall-normal location $y^+ = 15$ to sample the velocity. Figure 3 shows the distribution in the larger channel of the inner-scaled velocity-fluctuation components after the initial transient ($t^+ > 500$) in the streamwise (u) and wall-normal (v) directions. Opposition control yields a reduction of the intensity of the fluctuations in both directions when $y_s^+ = 10$ or $y_s^+ = 15$. The opposite effect is obtained when y_s^+ is further from the wall, leading to general increase both in the streamwise and wall-normal direction. When considering the DRL policies, a sensing plane close to the wall will lead to an increase of the wall-normal fluctuations and a reduction of the streamwise fluctuations, determining a significant modification of the fluctuations distribution. The changes are much less pronounced with $y_s^+ = 30$, where the DRL pol-

icy determines a slight reduction of intensity in both x and y direction, in a way that reminds the effect of opposition control with a sensing plane positioned closer to the wall. Finally, when $y_s^+ = 50$, the control has the opposite effect, with a slight increase of the fluctuations in both directions, which results in a small drag increase.

CONCLUSIONS

In this work we assess the effect of the location of the wall-parallel observation plane in both opposition control and in the DRL learning process. Both approaches yield drag reduction when the flow is sampled at $y_s^+ = 10$ and 15. In both these cases the policy obtained with DRL performs better than opposition control. Furthermore, if the sensing plane is located farther from the wall, the application of opposition control leads to significant drag increase. By contrast, the DRL algorithm is able to identify drag-reduction strategies also with observations at $y_s^+ = 30$ and $y_s^+ = 50$ in the minimal channel flow. We compare both the policies using input-output maps and the effect of the control on the velocity-fluctuation distribution in the flow.

When testing the policies in a larger channel flow, a performance reduction can be observed in all the considered cases. DRL agents trained in the minimal channel did not experience the entire range of physical features that are present in the larger channel, thus preventing their policies from providing equally good results in this latter setting. Sampling the flow at $y_s^+ = 30$ yields a limited drag reduction, whereas at $y_s^+ = 50$ we observe a drag increase. These results need to be improved to allow the described DRL approach to find application in all cases in which sensing the flow very close to the wall can be difficult. Different approaches can be used to achieve such goal, for instance transfer learning Pan & Yang (2009) or curriculum learning Bengio *et al.* (2009). Both leverage the fact that, even if the policies learnt in the minimal channel are not as effective in the larger channel, they represent a valid starting point for the training in the larger setting. The use of these techniques will be the subject of our future work.

ACKNOWLEDGEMENTS

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

REFERENCES

Bengio, Yoshua, Louradour, Jérôme, Collobert, Ronan & Weston, Jason 2009 Curriculum learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*, p. 41–48. New York, NY, USA: Association for Computing Machinery.

Biferale, Luca, Bonaccorso, Fabio, Buzzicotti, Michele, Clark Di Leoni, Patricio & Gustavsson, Kristian 2019 Zermelo's problem: optimal point-to-point navigation in 2D turbulent flows using RL. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **29** (10), 103138.

Choi, Haecheon, Moin, Parviz & Kim, John 1994 Active turbulence control for drag reduction in wall-bounded flows. *Journal of Fluid Mechanics* **262**, 75–110.

Font, Bernat, Alcántara-Ávila, Francisco, Rabault, Jean, Vinuesa, Ricardo & Lehmkuhl, Oriol 2024 Active flow control of a turbulent separation bubble through deep reinforcement learning. *Journal of Physics: Conference Series* arXiv:2403.20295.

Guastoni, Luca, Rabault, Jean, Schlatter, Philipp, Azizpour, Hossein & Vinuesa, Ricardo 2023 DRL for turbulent drag reduction in channel flows. *The European Physical Journal E* **46** (4), 27.

Jiménez, Javier & Moin, Parviz 1991 The minimal flow unit in near-wall turbulence. *Journal of Fluid Mechanics* **225**, 213–240.

Kurz, Marius, Offenhäuser, Philipp & Beck, Andrea 2023 DRL for turbulence modeling in large eddy simulations. *International Journal of Heat and Fluid Flow* **99**, 109094.

Lillicrap, Timothy P., Hunt, Jonathan J., Pritzel, Alexander, Heess, Nicolas, Erez, Tom, Tassa, Yuval, Silver, David & Wierstra, Daan 2015 Continuous control with DRL. *Preprint,* arXiv:1509.02971.

Lorsung, Cooper & Barati Farimani, Amir 2023 Mesh deep Q network: A DRL framework for improving meshes in computational fluid dynamics. *AIP Advances* **13** (1), 015026.

Novati, Guido, de Laroussilhe, Hugues Lascombes & Koumoutsakos, Petros 2021 Automating turbulence modeling by multi-agent reinforcement learning. *Nature Machine Intelligence* **3**, 87–96.

Pan, S. J. & Yang, Q. 2009 A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* **22** (10), 1345–1359.

Rabault, Jean, Kuchta, Miroslav, Jensen, Atle, Réglade, Ulysse & Cerardi, Nicolas 2019 Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *Journal of Fluid Mechanics* **865**, 281–302.

Sonoda, Takahiro, Liu, Zhuchen, Itoh, Toshitaka & Hasegawa, Yosuke 2023 RL of control strategies for reducing skin friction drag in a fully developed turbulent channel flow. *Journal of Fluid Mechanics* **960**, A30.

Suárez, Pol, Alcántara-Ávila, Francisco, Miró, Arnau, Rabault, Jean, Font, Bernat, Lehmkuhl, Oriol & Vinuesa, Ricardo 2023 Active flow control for three-dimensional cylinders through DRL. *Preprint,* arXiv:2309.02462.

Varela, Pau, Suárez, Pol, Alcántara-Ávila, Francisco, Miró, Arnau, Rabault, Jean, Font, Bernat, García-Cuevas, Luis Miguel, Lehmkuhl, Oriol & Vinuesa, Ricardo 2022 DRL for flow control exploits different physics for increasing Reynolds number regimes. *Actuators* **11** (12).

Verma, Siddhartha, Novati, Guido & Koumoutsakos, Petros 2018 Efficient collective swimming by harnessing vortices through DRL. *Proceedings of the National Academy of Sciences* **115** (23), 5849–5854.

Vignon, Colin, Rabault, Jean & Vinuesa, Ricardo 2023 Recent advances in applying deep reinforcement learning for flow control: Perspectives and future directions. *Physics of Fluids* **35** (3), 031301.

Wallace, James M., Eckelmann, Helmut & Brodkey, Robert S. 1972 The wall region in turbulent shear flow. *Journal of Fluid Mechanics* **54** (1), 39–48.

Zeng, Kevin, Linot, Alec J. & Graham, Michael D. 2022 Data-driven control of spatiotemporal chaos with reduced-order neural ODE-based models and RL. *Proceedings of the Royal Society A* **478** (2267), 20220297.

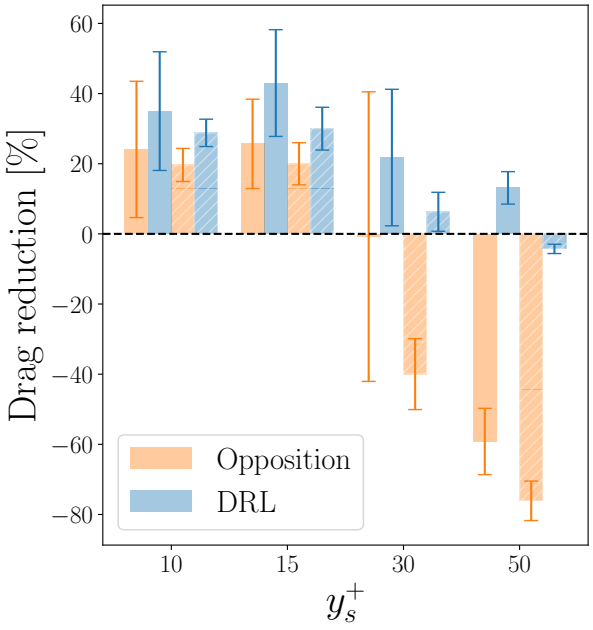

Figure 1. Average drag reduction with respect to the reference uncontrolled case using opposition or DRL control in the minimal channel and larger channel, as a function of the inner-scaled location of the sensing plane. The error bars represent the standard deviation of the drag reduction with respect to the initial conditions. Solid-color bars represent the results in the minimal channel, while the ones with white stripes denote the drag reduction in the larger channel.

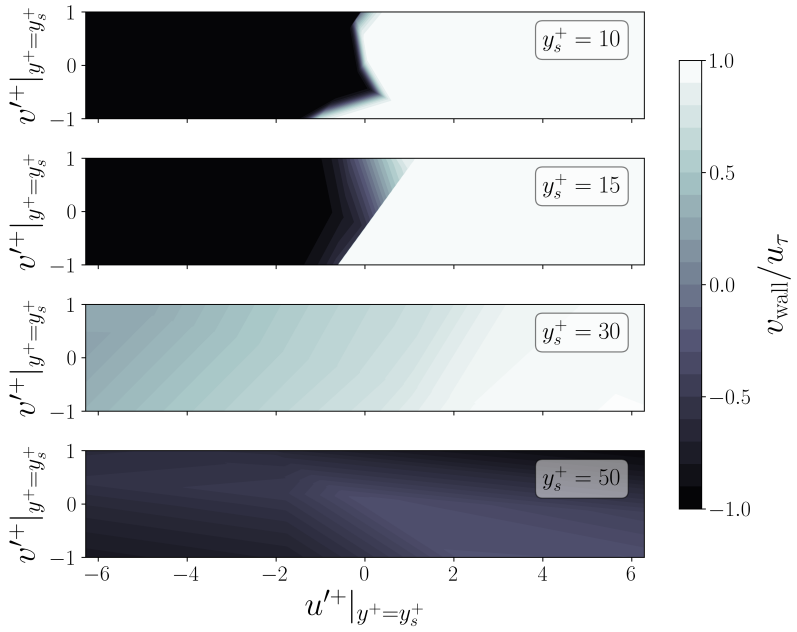

Figure 2. Input-output maps of the DRL control policies learnt in the minimal channel, using the state observed at the various sensing-plane locations y_s^+.