ISSN 2308-5916. Mathematical and computer modelling.
Series: Technical sciences. 2026. Issue 29. P. 95-107.

UDC 004.8; 004.9
DOI: 10.32626/2308-5916.2026-29.95-107

Petrovskyi O.
ORCID: 0000-0002-5729-544X
Ph.D. Student in Computer Science, Lviv Polytechnic
National University, Lviv, Ukraine,
E-mail: oleksandr.s.petrovskyi@lpnu.ua

Liaskovska S.
ORCID: 0000-0002-0822-0951
Ph.D. in Computer Science, Lviv Polytechnic
National University, Lviv, Ukraine,
E-mail: solomiya.y.lyaskovska@lpnu.ua

HYBRID FOURIER NEURAL OPERATOR BASED FRAMEWORK FOR ADAPTIVE CONTINUOUS CONTROL

Hybrid adaptive control methods are of high scientific interest and industrial urgency due to their ability to address the weaknesses of both model-driven and data-driven controllers, as the former are reliable and predictable, but rigid and often suboptimal, while the latter provide eventual optimality guarantees given enough time and exploration, but are effectively unapplicable to systems with a high cost of operation and failure. Thus, this work aims to develop and evaluate a hybrid framework for building reinforcement-learning-based adaptive controllers that leverages the unique properties of Fourier Neural Operators (FNOs) to achieve higher accuracy and reliability in control systems. To achieve this, we propose an architecture of a TD3-based agent that uses pretrained FNO as a world model. The framework includes training the FNO network on historical data, pretraining the agent on the obtained environment surrogate, adjusting the world model in real time based on new observations, and dynamically balancing between Dreamer-like planning and Q-network estimates depending on how well the world model can predict the system's response to the agent's actions. The effectiveness of the method was evaluated on the simulation of a baker’s yeast fermenter. Experimental results show that the proposed method significantly outperforms popular algorithms such as SAC, TD3, and TD3+CQL, as well as a pure FNO-based controller: the agent successfully reaches the target biomass con-

---

Стаття надійшла до редакції: 19.03.2026
Рекомендовано до друку: 07.04.2026
Оприлюднено (online): 15.05.2026

centration without risky exploration of the environment and demonstrates the ability to overcome model-reality mismatches, which proves its effectiveness and great potential. The practical value of this work lies in developing a method that enables the creation of reliable adaptive controllers for complex nonlinear processes with a high cost of failure, which do not require analytical models and can continually adjust themselves to real conditions, given the availability of historical data.

Key words: *adaptive control, hybrid reinforcement learning, neural operators, data-driven modeling, offline-to-online, pretraining, continuous control, machine learning, bioreactor.*

Introduction. Despite the popularity of reinforcement learning (RL) and many cases of its successful application to various control tasks, most of the large-scale industrial autonomous control systems are still implemented using traditional methods such as proportional integral derivative (PID) controllers or model predictive control (MPC). Such model-driven control methods allow for the utilization of prior knowledge and provide stable, reliable, and predictable operation. However, their drawback is rigidity – if there is a considerable model-reality mismatch, the control is guaranteed to be suboptimal, and the only way to address this issue is to redesign the model, which may be costly and require great expertise. On the other hand, data-driven methods allow for continuous adaptation by learning directly from empirical observations, so the policy will always be aligned with the actual environment and converge to the optimal control given enough time and exploration. However, many issues arise, namely, the fundamental exploration/exploitation trade-off, sample inefficiency, learning instability, poor generalization, catastrophic forgetting, overestimation of return, etc. This renders them inapplicable to environments with a high cost of operation and failure. To address limitations of both model-driven and data-driven controllers, hybrid methods are actively being researched, as properly balancing the two would allow for effective and reliable adaptive control.

Recent invention of Neural Operators, with the Fourier Neural Operator being one of the most popular, is reshaping the field of machine learning by providing a powerful tool for obtaining high-quality surrogates of complex continuous systems purely from data, that are also computationally efficient and fully differentiable. They already make it possible to achieve enormous speedups on massive scientific simulations, effectively solve optimal transport problems, and predict the evolution of highly complex systems in fluid dynamics, plasma physics, etc. Thus, this study aims to design a novel multistage framework for building hybrid reinforcement-learning-based adaptive controllers that fully leverage the advantages of Fourier Neural Operators, moving us closer to easily implementable, effective control of arbitrary systems of any complexity.

Related work.

Fourier Neural Operator. Neural Operator [1] is a novel neural network architecture using special differentiable transforms to project data into functional spaces (e.g., Banach space, frequency domain, space of wavelets, etc.), which allows for advanced interactions between input variables to be learned. This made them highly effective for many tasks, like accelerating scientific simulations [2-3], optimal transport [4], weather prediction [5], and many more. Fourier Neural Operator (FNO) [6] was found especially useful for learning solutions to Partial Differential Equation (PDE) systems, resulting in fully data-driven and highly efficient (for massive simulations, they can be 100-1000000 times faster than traditional solvers) target system surrogates that allow direct backpropagation and super-resolution [7]. Such properties make FNO a compelling method in the context of growing interest towards data-driven approximations of real-world systems.

When compared to other popular data-driven modelling methods, like SINDy [8], OpInf [9], and their modifications, which are interpretable and easier to use, training an accurate FNO has all the inherent complexities of neural networks and may not be the best choice for all cases. However, their flexibility and differentiability make them a great foundation for the model-based reinforcement learning because:

- Differentiable imaginary rollouts can be performed to allow Dreamer-like backpropagation through time for farsighted planning and computation of more accurate gradients.
- Since FNO predicts the next state of the system, we can generate synthetic data using it as an environment substitute or even pretrain controllers on it directly.
- By measuring prediction errors at every timestep, we can naturally detect data-drift and adjust the rollout horizon accordingly to prevent error amplification, choose safer actions, notify staff, etc.
- Continual learning methods can be applied to update the FNO-network on the fly to minimize the next state prediction error, as new evidence arrives, forming an Active-Inference-like mechanism of the world model refinement.

Apart from that, FNO is a great tool for modelling systems with primarily slow dynamics, like fermentation and other large-scale bioprocesses, as it has an embedded low-pass filter allowing it to capture the guiding rules of the environment while ignoring the measurement noise and small random disturbances.

The potential applicability and effectiveness of FNO in adaptive control systems are actively being researched. In [10] Dutta et al present an FNO-based encoder for model-free visual reinforcement learning that provides a considerable performance gain over traditional CNN encoders and validates this approach across popular benchmarks. Bhan et al [11] developed an adaptive control scheme augmented with neural operators that demonstrated 98% faster control with high accuracy in 2 non-linear feedback control systems while preserving the closed-loop stability guarantees. Lin et al [12] developed a data-driven Runge-Kutta scheme for DeepONet that allows for medium/long-term system response prediction and validated it on a set of controlled PDE-driven systems, including Cart Pole. Fourier Q Operator Network (FQON) architecture, implemented in [13] and successfully validated in an industrial setting, learns direct mapping between expected state function and Q-function to efficiently achieve implicit discretization-invariant Q-value estimates. Additionally, a well-trained FNO network can also be used outside of the controlling algorithm for offline pretraining or augmentation of the RL agent’s replay buffer with synthetic data in real time, as it is a full-fledged environment surrogate.

Hybrid Reinforcement Learning. Many methods for combining model-free and model-based control exist, each having its strengths and weaknesses. Apart from well-established general approaches to utilizing world models, like 3 generations of Dreamer [14], learning-based MPC [15], model-based value expansion [16] and other [17], more specific methods are developed to take advantage of certain features of a concrete control objective. For example, in [18] authors implement and validate a hybrid Dyna-style controller combining SINDy with TD3 to greatly reduce the exploration needs of the latter in environments with dynamics decomposable into a set of interacting mathematical functions. This is achieved by building an approximation of the target system with SINDy and using it to augment the agent’s memory replay buffer with synthetic data. The dynamics model is first trained in an offline manner, and, if needed, adjusted on the fly based on online observations. Another powerful method is ROL [19], which builds on a residual learning framework by utilizing an offline-trained DeepONet (also a neural operator) to capture the general system’s dynamics and select the best action, while an online TD3 agent learns to mitigate the mismatches between the real system and its model. Some works take inspiration from nature, mimicking the way humans arbitrate between fast, reflexive model-free and slow, thoughtful model-based decision-making [20], such as [21], that describes value-of-information-based arbitration, which utilizes model-based planning in situations of uncertainty, otherwise sticking to regular Q-learning. It has been demonstrated that, given the policies arbitrated between are diverse enough, significant performance improvements can be achieved in the continual learning setting [22].

In the proposed framework, we incorporate successful ideas described above by also utilizing model-based (FNO) and model-free (TD3) branches, which are trained simultaneously, and arbitrating between them based on the world model's ability to correctly predict the next state. When the world model is accurate, more weight is given to farsighted planning via the Dreamer-like imaginary environment rollouts, but when FNO struggles to predict the next state, indicating a data drift, a regular TD3 Critic comes into play as a fallback, until FNO captures new dynamics. In comparison with SINDy-TD3, using FNO allows for more natural integration of the world model into an agent (backpropagation through time and environment vs replay buffer extension) and immediate high-quality control, as FNO is also used for pretraining, while, in contrast with ROL, we actually handle the model-reality mismatch with new observations becoming available by means of continual learning.

Proposed Approach. The overall agent's architecture is based on Twin Delayed Deep Deterministic Policy Gradient (TD3) [23] augmented with an FNO network, and consists of such components: *Actor* network $\pi_\theta : S \to A$, selecting an action based on the observation, two *Critic* networks $Q_{\varphi_i} : S \times A \to \mathbb{R},\ i \in \{1,2\}$, evaluating the decision of the Actor, and a *World* network $W : \vec{S} \times A \to S$, predicting the next state of the system s_{t+1} based on $h \in \mathbb{N}$ recent observations $s_{t-h}, s_{t-h+1} \ldots s_t$ and the action taken a_t. Original TD3 Actor's loss is augmented with a backpropagation-through-time loss term $\mathcal{L}_{BTT}$ weighted by the trust parameter $\lambda \in (0,1)$ as shown in (1).

$$-\mathcal{L}_\pi = (1-\lambda)\cdot Q(s,a) + \lambda \cdot \mathcal{L}_{BTT}. \tag{1}$$

$\mathcal{L}_{BTT}$ is calculated by computing the cumulative reward over a Dreamer-like imaginary environment rollout of length K using the heuristic reward function $r : S \to \mathbb{R}$ and weighting each result by the decay coefficient γ. If we denote the prediction of W at the k-th rollout as $\hat{s}_k = W(\hat{s}_{k-1}, \hat{a}_{k-1})$, with $\hat{a}_k = \pi(\hat{s}_k)$, $\hat{a}_0 = \pi(s)$, and $\hat{s}_0 = s$, the calculation of $\mathcal{L}_{BTT}$ can be represented as (2).

$$\mathcal{L}_{BTT} = r(s) + \sum_{k=1}^{K} \gamma^k \cdot r(\hat{s}_k). \tag{2}$$

Total mechanism of Actor gradient computation is illustrated in Fig.1.

Such a combined loss will make the Actor maximize both the Q-value, which would likely be localized to a current region of the state space (performance based on empirical data), and the imaginary loss incorporating the general rules of the system (model-based performance).

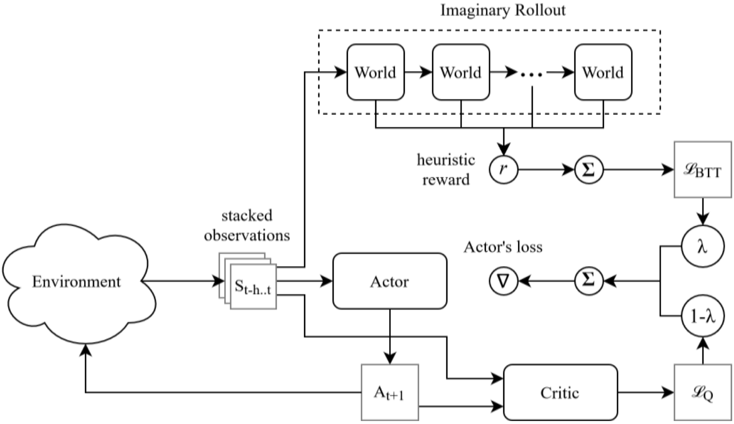

Fig. 1. *Hybrid Actor-network update mechanism*

As World-network W was trained to predict state transitions, we can detect data drifts in real time by comparing the predicted state $\hat{s}_t = W\left(s_{t-1}, a_{t-1}\right)$ with the observed state s_t at every timestep. If the prediction error is high, we lower the trust parameter λ making the agent favor the model-free branch more and vice versa, as shown in (3), avoiding the excessive error amplification and giving time for W to capture the unknown mechanics.

$$\lambda_{t+1} = \lambda_{min} + \left(\lambda_{max} - \lambda_{min}\right) \cdot e^{-\beta \cdot \max\left(\epsilon_t, 0\right)}, \tag{3}$$

where $\epsilon_t = \dfrac{s_t - \hat{s}_{t2}}{\mathrm{EMA}\left(\epsilon\right)}$, is a world model prediction error normalized by the exponential moving average of the error magnitude ϵ, $\lambda_{\max}$ and $\lambda_{\min}$ are maximum and minimum values the trust parameter λ can take (we used $\lambda_{\max} = 0.99$ and $\lambda_{\min} = 0.01$), and β is a hyperparameter determining the sensitivity of λ to the prediction error $s_t - \hat{s}_{t2}$.

Before integrating the agent of such complexity into the system of interest, we warm it up by pretraining it for several epochs on a copy of W, using the reward function r to compute the value of every state s_{t+1}.

Experimental setup. Industrial baker's yeast fermentation process was chosen as a control objective for this study as it possesses several properties making both purely model-driven and data-driven methods suboptimal, namely *high complexity* (making precise analytical models hard to build, while RL-based methods often struggle to learn a good policy due to the non-linearity, stochasticity, etc.), *delayed response* (an effect of change in feed rate at the current timestep may only become noticeable many hours later, making it quite hard for an RL-based controller to associate the action taken with the specific outcome and differentiate it from outcomes of all other decisions made during that period), *variability and non-stationarity* (each bioreactor is different, the course of fermentation process varies even within the same device from batch to batch, and many hard-to-predict factors such as byproduct accumulation or sediment buildup on sensors may negatively affect the applicability of the model), *high operation and failure cost* (it is economically infeasible to perform hundreds or thousands of training episodes with real fermenters just for the agent to explore the environment repeatedly failing the yield), *slow operation* (a single batch may take a week to complete, so even if running 1000 batches was possible, it would take 27 years to train the agent) and so on.

For experiments, we have chosen a simulation based on the analytical model of baker yeast fermentation process (5) used by Pandian et al. in [24], as it is based on the Monod equation of growth of microorganisms with strong empirical support, and, despite its simplicity, it produces a non-linear dynamics with delayed responses. The meaning, domain and initial values of variables and parameters of the simulation are given in Table 1.

$$\begin{gathered}\frac{dx_1}{dt} = \left(r - u_1 - \theta_4\right) \cdot x_1, \\ \frac{dx_2}{dt} = -\frac{rx_1}{\theta_3} + u_1 \cdot \left(u_2 - x_2\right), \\ r = \frac{\theta_1 x_2}{\theta_2 + x_2}.\end{gathered} \tag{5}$$

A small noise $\epsilon \in \mathcal{N}\left(0, \sigma_{\mathcal{O}}\right)$ was added to each state variable $x_{1,2}$ to better reflect real conditions with unavoidable deviations and measurement errors, where $\mathcal{N}\left(\mu, \sigma\right)$ is a normal distribution with a mean of μ and the standard deviation of σ.

Table 1

Parameters of the bioreactor simulation

Symbol	Meaning	Value
x_1	Biomass concentration (g/l)	7.0 (initial)
x_2	Substrate concentration (g/l)	0.1 (initial)
u_1	Dilution factor (h^{-1})	0.1
u_2	Substrate in the feed (g/l)	action, $u_2 \in [5,35]$
$\theta_{1..4}$	Other process parameters	$0.31, 0.18, 0.55, 0.05$
x_1^*	Desired biomass concentration (g/l)	10.0
$\sigma_{\mathcal{O}}$	Observation noise standard deviation (g/l)	0.05

The dataset representing historical observations was generated by running the environment 100 times for 1000 timesteps with a random control policy – choosing random action from the action space at every timestep.

Results and discussion. The proposed method's performance was compared against SAC [25] serving as a model-free baseline, original TD3 [23] as an ablation study, CQL [26] as a popular technique for offline pretraining (in this case, of TD3), and a pure FNO-based controller to assess the ability to adapt to model-reality mismatch. All agents were run for 1000 timesteps of a single episode (to imitate the deployment to a real system). Dynamics of biomass concentration and actions taken (Fig. 2), as well as rewards and a slice of a learned policy across a fixed $x_2 = 0.5$ at the last timestep (Fig. 3), were recorded for each agent. The sum, mean, and time-weighted cumulative errors are listed in Table 2.

Without extra tricks like reward engineering, TD3 was repeatedly failing to converge with all tried hyperparameters (even after 100 episodes) due to vanishing gradients that forced the policy to be stuck at the maximum possible action, which is beneficial at the beginning, but harmful as biomass concentration grows. Preinitialized networks obtained via offline pretraining with the CQL algorithm allowed TD3 to eventually converge in 20 episodes, but this wasn't without being stuck at the maximum action again for most of the evaluation episode. This is explained by CQL not being able to produce such initialization of a Critic network that successfully handles states not seen in historical observations, forcing TD3 to explore the state space and get stuck. SAC was able to promptly figure out the rules of the environment and drive the system towards the desired state; however, both the control and eventual policy were far from safe and stable, as seen in Figures 2 & 3. Only our approach and pure FNO-controller were able to quickly reach the target biomass concentration of 10 g/l and reliably maintain it for the evaluation period. Our method managed to improve a pure FNO-based performance.

Experiment results demonstrate several advantages of our framework over other evaluated methods. Pretraining allows for decreasing exploration needs and performing safer and more efficient control from the very start.

Table 2

Rewards of agents (less is better) across 1000 timesteps of operation

Metric	**Ours**	**FNO**	**SAC**	**TD3**	**TD3 + CQL**
Sum	-6.898	-7.513	-107.684	-2746.430	-2744.666
MAE	0.007	0.008	0.108	2.746	2.745
MSE	0.011	0.011	0.027	7.581	7.571
ITAE	365.694	673.979	51380.29	1384132.19	1383277.00
ITSE	6.854	7.504	8101.893	3835702.43	3830966.06

Using data-driven modeling instead of plain model-free pretraining allows for a greater amount of available knowledge to be captured, as simulations can accurately extrapolate to small regions around known data points, which is especially beneficial in the context of bioprocess control, where the system's state primarily changes slowly, making it easily decomposable into low-frequency dynamics with greater regions of accurate extrapolation. Of course, this puts additional requirements on the well-trainedness of FNO, as every error is amplified during pretraining, but, again, operating in low frequencies helps to mitigate the noise and small random disturbances, which makes FNO a great choice for similar systems. Incorporating the Dreamer-like imaginary rollouts made it possible to overcome the vanishing gradients issue observed in pure TD3, helping it to converge to a good initial policy. The method also showed the ability to improve with more observations.

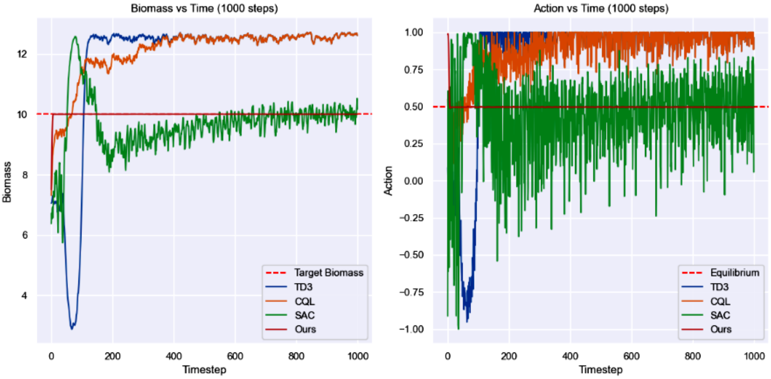

Fig. 2. *Visualization of biomass dynamics (left) and actions taken by the adaptive controller at each timestep (right)*

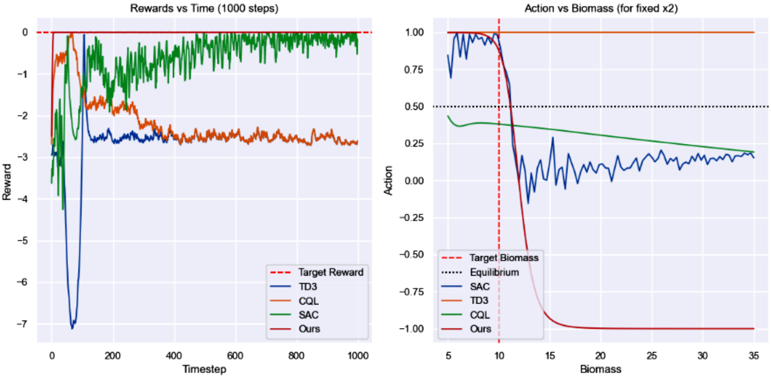

Fig. 3. *Visualization of reward dynamics (left) and agents' control policies at 1000-th timestep for varying* x_1 *and fixed* $x_2 = 0.5$ *(right)*

However, several issues were also observed. Computational complexity and the speed of weight update increased dramatically due to imaginary rollouts. FNOs were shown to be faster than numerical solvers for big simulations, but on small scales, the time to compute several forward passes may render the method unapplicable for environments requiring immediate decisions. Also, the arbitration mechanism, where the agent alternates between model-based and model-free branches during policy updates based on λ, was found to be less stable than expected: sometimes model-free-favoring updates temporarily decreased the agent's performance, which suggests that a more stable method of combining model-based and model-free branches needs to be developed. Apart from that, the evaluation environment was fully observable, creating a need for additional experiments in partially observable settings.

Conclusions. This paper designs a novel multistage method aiming to fully utilize all benefits of FNO in the context of RL-based adaptive continuous control such as ability to construct an accurate system surrogate from historical observations (especially effective for PDE-based systems with slow dynamics due to the built-in low-pass filter), perform differentiable Dreamer-like environment rollouts to aid the convergence to a good policy, detect data-drifts by evaluating the prediction error and take actions to mitigate the error amplification, continually update the world model with new evidence in an Active Inference fashion, which in combination produces a solid framework for implementing a safe and reliable adaptive control. The approach was evaluated against popular algorithms, showing a significant advantage both in safety and accuracy of the control. Also, current limitations of the approach were discussed and will be addressed in future work.

References:

1. Kovachki N. et al. Neural Operator: Learning Maps Between Function Spaces. May 02, 2024. DOI: 10.48550/arXiv.2108.08481.
2. Pamela S. J. P. et al. Neural-Parareal: Self-improving acceleration of fusion MHD simulations using time-parallelisation and neural operators. *Computer Physics Communications.* 2025. Vol. 307. P. 109391. DOI: 10.1016/j.cpc.2024.109391.
3. Azizzadenesheli K., Kovachki N., Li Z. et al. Neural operators for accelerating scientific simulations and design. *Nat Rev Phys.* 2024. Vol. 6. № 5. P. 320-328. DOI: 10.1038/s42254-024-00712-5.
4. Geuter J., Kornhardt G., Tomasson I., Laschos V. Universal Neural Optimal Transport. *Proceedings of the 42nd International Conference on Machine Learning,* PMLR, Oct. 2025. P. 19196-19232. URL: https://proceedings.mlr.press/-v267/geuter25a.html (Accessed: Mar. 13, 2026.)
5. Kurth T. et al. FourCastNet: Accelerating Global High-Resolution Weather Forecasting Using Adaptive Fourier Neural Operators. *Proceedings of the*

Platform for Advanced Scientific Computing Conference. PASC ’23. New York, NY, USA: Association for Computing Machinery, 2023. P. 1-11. DOI: 10.1145/3592979.3593412.

6. Li Z. et al. Fourier Neural Operator for Parametric Partial Differential Equations. May 17, 2021. *arXiv*: arXiv:2010.08895. DOI: 10.48550/arXiv.2010.08895.
7. Duruisseaux V., Kossaifi J., Anandkumar A. Fourier Neural Operators Explained: A Practical Perspective. Jan. 22, 2026. *arXiv*: arXiv:2512.01421. DOI: 10.48550/arXiv.2512.01421.
8. Brunton S. L., Proctor J. L., Kutz J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*. 2016. Vol. 113. № 15. P. 3932-3937. DOI: 10.1073/pnas.1517384113.
9. Peherstorfer B., Willcox K. Data-driven operator inference for nonintrusive projection-based model reduction. *Computer Methods in Applied Mechanics and Engineering*. 2016. Vol. 306. P. 196-215. DOI: 10.1016/j.cma.2016.03.025.
10. Dutta P., Ayyoob M., Bhatnagar S., Dukkipati A. One Encoder to Rule them All: Representation Learning for Model-free Visual Reinforcement Learning using Fourier Neural Operators. *Presented at the Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2025. P. 4818-4827. URL: https://openaccess.thecvf.com/content/ICCV2025/html/Dutta_One_Encoder_to_Rule_them_All_Representation_Learning_for_Model-free_ICCV_-2025_paper.html (accessed: Mar. 12, 2026).
11. Bhan L., Shi Y., Krstic M. Operator Learning for Nonlinear Adaptive Control. *Proceedings of The 5th Annual Learning for Dynamics and Control Conference*. PMLR. 2023. P. 346-357. URL: https://proceedings.mlr.press/v211/-bhan23a.html (Accessed: Mar. 13, 2026).
12. Lin G., Moya C., Zhang Z. On Learning the Dynamical Response of Nonlinear Control Systems with Deep Operator Networks. Sep. 26, 2023. *arXiv*: arXiv:2206.06536. DOI: 10.48550/arXiv.2206.06536.
13. Ni Y., Li Y., Liu C., Jin Y. A New Fourier Q Operator Network Based Reinforcement Learning Method for Continuous Action Space Decision-making in Manufacturing. *Robotics and Computer-Integrated Manufacturing*. Vol. 86. P. 102641. Apr. 2024. DOI: 10.1016/j.rcim.2023.102641.
14. Hafner D., Pasukonis J., Ba J., Lillicrap T. Mastering diverse control tasks through world models. *Nature*. 2025. Vol. 640. № 8059. P. 647-653. DOI: 10.1038/s41586-025-08744-2.
15. Hewing L., Wabersich K. P., Menner M., Zeilinger M. N. Learning-Based Model Predictive Control: Toward Safe Learning in Control. *Annual Review of Control, Robotics, and Autonomous Systems*. 2020. Vol. 3. № 3. P. 269-296. DOI: 10.1146/annurev-control-090419-075625.
16. Feinberg V., Wan A., Stoica I., Jordan M. I., Gonzalez J. E., Levine S. Model-Based Value Estimation for Efficient Model-Free Reinforcement Learning. Feb. 28, 2018, *arXiv*: arXiv:1803.00101. DOI: 10.48550/arXiv.1803.00101.
17. Luo F.-M., Xu T., Lai H., Chen X.-H., Zhang W., Yu Y. A Survey on Model-based Reinforcement Learning. Jun. 19, 2022. *arXiv*: arXiv:2206.09328. DOI: 10.48550/arXiv.2206.09328.

18. Abdelsalam K., Gamal Z., El-Badawy A. Dyna-Style Reinforcement Learning Modeling and Control of Non-linear Dynamics. Dec. 24, 2025. *arXiv*: arXiv:2512.21081. DOI: 10.48550/arXiv.2512.21081.
19. Ahmed N. et al. Reinforcement Operator Learning (ROL): A hybrid DeepONet-guided reinforcement learning framework for stabilizing the Kuramoto–Sivashinsky equation. *PLOS ONE*. 2026. Vol. 21. № 1. P. e0341161. DOI: 10.1371/journal.pone.0341161.
20. Kool W., Gershman S. J., Cushman F. A. Cost-Benefit Arbitration Between Multiple Reinforcement-Learning Systems. *Psychol Sci*. 2017. Vol. 28. № 9. P. 1321-1333. DOI: 10.1177/0956797617708288.
21. Bera K., Mandilwar Y., Raju B. Value-of-Information based Arbitration between Model-based and Model-free Control. Dec. 08, 2019. *arXiv*: arXiv:1912.05453. DOI: 10.48550/arXiv.1912.05453.
22. Jang W., Kim D. An Arbitration Control for an Ensemble of Diversified DQN variants in Continual Reinforcement Learning. Sep. 05, 2025. *arXiv*: arXiv:2509.04815. DOI: 10.48550/arXiv.2509.04815.
23. Fujimoto S., Hoof H. van, Meger D. Addressing Function Approximation Error in Actor-Critic Methods. Oct. 22, 2018. *arXiv*: arXiv:1802.09477. DOI: 10.48550/arXiv.1802.09477.
24. Pandian B. J., Noel M. M. Control of a bioreactor using a new partially supervised reinforcement learning algorithm. *Journal of Process Control*. 2018. Vol. 69. P. 16-29. DOI: 10.1016/j.jprocont.2018.07.013.
25. Haarnoja T., Zhou A., Abbeel P., Levine S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. Aug. 08, 2018. *arXiv*: arXiv:1801.01290. DOI: 10.48550/arXiv.1801.01290.
26. Kumar A., Zhou A., Tucker G., Levine S. Conservative Q-learning for offline reinforcement learning. *Proceedings of the 34th International Conference on Neural Information Processing Systems*. NIPS ’20. Red Hook, NY, USA: Curran Associates Inc., Dec. 2020. P. 1179-1191. URL: https://dl.acm.org/-doi/10.5555/3495724.3495824 (accessed: Mar. 13, 2026).

ГІБРИДНИЙ ФРЕЙМВОРК ДЛЯ АДАПТИВНОГО НЕПЕРЕРВНОГО КОНТРОЛЮ НА БАЗІ НЕЙРОННИХ ОПЕРАТОРІВ ФУР’Є

На методи гібридного адаптивного контролю спрямований неабиякий науковий та промисловий інтерес через їх потенційну можливість поєднати в собі сильні сторони методів контролю на базі аналітичних моделей (model-driven) та емпіричних даних (data-driven), перші з яких надійні та передбачувані, але повністю позбавлені адаптивності, а останні гарантують оптимальний контроль, але через суттєві обмеження є фактично незастосовними у реальних виробничих умовах з високими ризиками і вимогами до якості. Метою цієї роботи є розробка та апробація гібридного фреймворку для побудови адаптивних контролерів на базі навчання з підкріпленням, який використовує особливості нейронних операторів Фур’є (FNO) для забезпечення вищої точності і надійності системи керування. Для досягнення цієї мети запропоновано архітектуру агенту, що базується на алгоритмі TD3 та використовує FNO у ролі моделі світу. Фреймворк включає тренування FNO-мережі на історичних даних, попереднє навчання агенту на отриманому сурогаті середовища, оновлення моделі світу на базі нових спостережень у реальному часі та динамічне балансування між Dreamer-подібним плануванням та оцінками Q-мереж в залежності від того, наскільки добре модель світу здатна передбачати реакцію системи на дії агенту. Ефективність методу оцінено на симуляції промислового процесу ферментації хлібопекарських дріжджів. Результати експериментів демонструють, що запропонований метод значно перевершує поширені алгоритми SAC, TD3 та TD3+CQL, а також чистий FNO-контролер: агент успішно досягає цільової концентрації біомаси без ризикованого дослідження середовища та демонструє здатність долати невідповідність моделі реальності. Практична цінність роботи полягає у розробці методу, за допомогою якого за наявності історичних даних можливо створювати надійні адаптивні контролери для складних нелінійних процесів з високою вартістю помилки, що не вимагають аналітичних моделей, та здатні самостійно підлаштовуватися під реальні умови.

Ключові слова: *адаптивний контроль, гібридне навчання з підкріпленням, нейронні оператори, моделювання на базі даних, офлайн-онлайн навчання, претренування, неперервний контроль, машинне навчання, біореактор.*