Proceedings of Machine Learning Research vol 242:1–13, 2024

In Vivo Learning-Based Control of Microbial Populations Density in Bioreactors

Sara Maria Brancato SARAMARIA.BRANCATO@UNINA.IT
University of Naples Federico II

Davide Salzano DAVIDE.SALZANO@UNINA.IT
Scuola Superiore Meridionale

Francesco De Lellis FRANCESCO.DELELLIS@UNINA.IT
University of Naples Federico II

Davide Fiore DAVIDE.FIORE@UNINA.IT
University of Naples Federico II

Giovanni Russo* GIOVARUSSO@UNISA.IT
University of Salerno

Mario di Bernardo* MARIO.DIBERNARDO@UNINA.IT
University of Naples Federico II

*Corresponding authors

Abstract

A key problem in using microorganisms as bio-factories is achieving and maintaining cellular communities at the desired density and composition to efficiently convert their biomass into useful compounds. Bioreactors are promising technological platforms for the real-time, scalable control of cellular density. In this work, we developed a learning-based strategy to expand the range of available control algorithms capable of regulating the density of a single bacterial population in bioreactors. Specifically, we used a *sim-to-real* paradigm, where a simple mathematical model, calibrated using a single experiment, was adopted to generate synthetic data for training the controller. The resulting policy was then exhaustively tested *in vivo* using a low-cost bioreactor known as Chi.Bio, assessing performance and robustness. Additionally, we compared the performance with more traditional controllers (namely, a PI and an MPC), confirming that the learning-based controller exhibits similar performance *in vivo*. Our work demonstrates the viability of learning-based strategies for controlling cellular density in bioreactors, making a step forward toward their use in controlling the composition of microbial consortia.

Keywords: Control Applications, Learning-Based Control, In Vivo Validation, Sim-To-Real, Synthetic Biology

1. Introduction

Microorganisms, such as bacteria and yeast, have been widely used in industry as efficient, low-waste bio-factories capable of converting nutrients into useful proteins or chemicals (Brenner et al., 2008; Satyanarayana, 2009; Su et al., 2020; Jullesson et al., 2015; Choi et al., 2018; Hug et al., 2020). These bio-factories offer significant benefits, including sustainability and reduced environmental impact, by leveraging biological processes for production. This is achieved by engineering *de novo* synthetic circuits into cells or combining the natural bio-processing capabilities of different organisms.

© 2024 S.M. Brancato, D. Salzano, F. De Lellis, D. Fiore, G. Russo* & M. di Bernardo*.

In this context, an important issue is how to utilize cell resources to efficiently transform biomass into protein production while preventing the accumulation of toxic by-products (Mauri et al., 2020; Tian et al., 2020; Xu et al., 2018; Lv et al., 2019). Bioreactors play a crucial role in this process. They provide a controlled environment where it is possible to achieve and maintain a desired cell density, creating optimal conditions for the bio-production of a given chemical. Figure 1 presents an example of automated control architecture applied to cell growth regulation. Specifically, by modulating the dilution with the introduction of new nutrients, it is possible to adjust the culture density in *real time*. An external controller can be designed to run on a computer and automatically regulate the cell density by evaluating the error between the measured density inside the chamber and the desired density level.

Various strategies exist for regulating cell populations within a chamber, including those that manipulate dilution rates in chemostats (De Leenheer and Smith, 2003) and those that leverage genetic interventions of cell strains, utilizing different control inputs such as light (Gutiérrez Mena et al., 2022; Lugagne et al., 2024) or various nutrients (Treloar et al., 2020). From a control design perspective, existing approaches utilize traditional controllers like PIs (Kusuda et al., 2021), non-linear piecewise smooth methods, or gain scheduling state feedback strategies (Fiore et al., 2021). Some approaches harness computational capabilities to derive control laws incorporating constraints, either through mechanistic models (Bertaux et al., 2022; Aditya et al., 2021; Zhu et al., 2000) or through data-driven methods utilizing reinforcement learning (Treloar et al., 2020) or deep neural networks (Lugagne et al., 2024).

Recent developments in quantitative systems and synthetic biology have led to the increased adoption of compact and cost-effective bioreactors, such as those explored by Bertaux et al. (2022); Steel et al. (2020); Wong et al. (2018). These bioreactors offer integrated control equipment and multiple sensors in a unified platform, enabling precise manipulation of environmental conditions for extended periods in microbial cultures, making them highly attractive for controlling microbial consortia. Among the various low-cost, open-source bioreactor platforms available for the rapid prototyping of novel microbial communities for bio-production, the Chi.Bio (Steel et al., 2020) provides a controlled, static environment where culture parameters such as nutrient availability and temperature can be regulated. It also includes the ability to frequently measure cellular density and bulk fluorescence, and offers optogenetic actuation. This platform employs a PI controller for the *real time* control of cell density in the culture vial. However, to enhance robustness, it is common practice to optimally tune the control gains using a mathematical model of the controlled system (Fiore et al., 2021; Wong et al., 2018). While being effective, this often requires accurate and well-calibrated mathematical models, which can be challenging to obtain.

To overcome the need for an accurate mathematical model, several approaches have resorted to data-driven modeling. In (Lugagne et al., 2024), the authors trained a deep network to accurately predict the expression of single cells in a microfluidic device. They used light stimuli to induce or not the expression of a fluorescence protein. For each cell in the device, they collected various features such as fluorescence, cell count in each chamber, and light inputs. The dataset included 16.000 time series approximately, which enabled extensive training of the model. Whereas the architecture of the neural network comprises two components: a long short-term memory that encodes the data series of each feature into a unique lower-dimensional vector, and a multi-layer perceptron that predicts fluorescence evolution. This deep network was then used as a model in a Model Predictive Control scheme to regulate the expression of thousands of cells based on their response to light stimuli. However, applying this approach to low-cost bioreactors like Chi.Bio is challenging, if

not impossible, because data from these bioreactors are collected from aggregate populations rather than individual cells, making data efficiency a significant issue. Another strategy is to employ a learning-based control approach, where data are used to directly learn the controller rather than the model. As proposed by Treloar et al. (2020), a suitable control law for a fixed reference can be learned within 24 hours using five parallel bioreactors. However, this approach lacks *in vivo* validation, where noise and population heterogeneity could hinder controller performance. Additionally, if the reference point changes, entirely new training is required to adjust the system to the new set point due to difficulties in training and on how the policy is shaped (Zhang et al., 2023; Zhao et al., 2022; Hafner and Riedmiller, 2011). In practice, this approach inherits the typical challenges of reinforcement learning, such as sample inefficiency, which requires extensive time and a large amount of experimental data (see Buşoniu et al. (2018); Bertsekas (2005)).

A possible solution to learn a control policy without requiring extensive experimental data is the *sim-to-real* approach, where the control policy is learned in simulated environments and subsequently transferred to the real system (Rusu et al., 2017; Tan et al., 2018; James et al., 2017). This is particularly challenging in biological applications, as these systems evolve and grow, characterized by cell-to-cell variability, uncertainties, and other disturbances that are difficult to accurately capture in synthetic mathematical models. Therefore, a key open problem is to determine if and how learning-based controllers trained with a simple model parametrized on limited experimental data can be effectively deployed *in vivo* to control bacterial populations, bridging the gap between simulations and real-world experiments.

In this work, we address this problem by developing a learning-based controller for regulating cellular density to different desired set points in a bioreactor. In line with the *sim-to-real* approach, the control law is learned by interacting with synthetically generated data. These data are generated from a simple deterministic model capturing the main features of the growth dynamics. Notably, even though partial knowledge of the system's dynamics is required, a coarse calibration of the parameters, obtained using a few open-loop experiments, is sufficient to generate the data needed for training the control algorithm to achieve set point regulation. We demonstrate through a series of exhaustive *in vivo* experiments that the *sim-to-real* gap can be bridged and that the control performance learned using a simple model can be transferred to real experiments conducted using a bioreactor. We benchmark our controller in terms of performance and robustness against the on-board PI controller in the Chi.Bio and a Model Predictive Controller developed for comparison, which employs the same simple deterministic model used to generate synthetic data for the learning-based controller.

2. Control Problem Formulation

We consider the growth dynamics of a bacterial species inside a microbial culture chamber described as a continuous-time dynamical system of the form:

$$\begin{aligned} \dot{x}_t &= f(x_t, u_t), \quad x_0 = \tilde{x}_0, \\ y_t &= \alpha\, x_t, \end{aligned} \tag{1}$$

where $x_t \in \mathcal{X}$ is the concentration of bacteria in the microbial culture chamber at time t, with $\mathcal{X} \subseteq \mathbb{R}_{\geq 0}$ being the state space, $\tilde{x}_0 \in \mathcal{X}$ is the initial concentration, $u_t \in \mathcal{U}$ is the control input or *pump rate* delivered as an exogenous injection of fresh growth media in the microbial chamber, with $\mathcal{U} \subseteq \mathbb{R}_{\geq 0}$ being the set of feasible inputs, $f : \mathcal{X} \times \mathcal{U} \to \mathcal{X}$ is the vector field defining the system

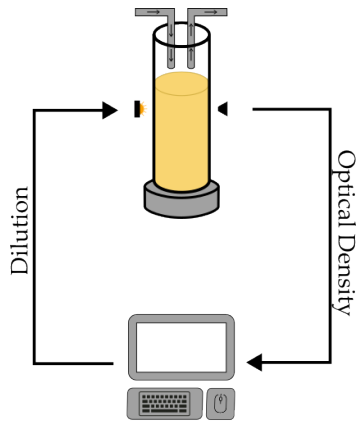

Figure 1: An automated cell growth setup: the cell density at a given time is estimated via optical density measures, while a computer automatically implements a control law able to regulate the dilution (hence the density inside the chamber) by adding fresh media and discarding the waste.

dynamics, and the output $y \in [0, 1]$ is the optical density (OD) measured by the platform, expressed in arbitrary units. For the sake of simplicity, we assume that α is equal to 1, and therefore from now on we will equivalently refer to either bacterial concentration x or optical density y. This causes only a simple rescaling of the output signal, and does not affect the control design that follows.

To accommodate the technological constraints of common microbiology platforms, we consider a scenario where the control input can only be applied at fixed discrete time steps. Therefore, we design our control strategy based on the following discrete-time dynamical system:

$$x_{t_{k+1}} = x_{t_k} + \int_{t_k}^{t_{k+1}} f(x_\tau, u_{t_k}), d\tau, \quad x_{t_0} = \tilde{x}_0, \tag{2}$$

where $t_k \in \mathbb{N}_{\geq 0}$ represent discrete time steps, and u_{t_k} is a piecewise constant function defining the constant *pump rate* applied during the time interval $[t_k, t_{k+1})$ to the system dynamics described in (1). Moreover, when $u_{t_k} \neq 0$, indicating that fresh media is being pumped into the chamber, the experimental platform automatically expels some of the fluid from the chamber at a rate greater than the input rate to prevent overflow.

Considering the following assumptions:

A1. The concentration x is quantified through OD measurements,

A2. The measures are collected at every minute,

A3. The control input, i.e. the pumping rate, is limited to avoid overflows,

the *control goal* is to regulate the bacterial concentration x in the chamber to some desired steady-state value $\bar{x} \in [0.2, 1]$. This range corresponds to the operating conditions where cells are in the exponential growth phase, which facilitates protein production.

2.1. Stating the Learning-Based Control Problem

Following (Recht, 2019), the previous control goal can be reformulated as a learning-based control problem. Specifically, we aim to learn the control policy $\pi : \mathcal{X} \to \mathcal{U}$ to solve the following dynamic

optimization problem over a finite time horizon $t_N \in \mathbb{N}_{>0}$:

$$\max_{\pi} \; J^{\pi}, \tag{3a}$$

$$\text{s.t.} \; x_{t_{k+1}} = x_{t_k} + \int_{t_k}^{t_{k+1}} f(x_\tau, u_{t_k})\, d\tau, \quad t_k \in \{0, \dots, t_{N-1}\}, \tag{3b}$$

$$u_{t_k} = \pi(x_{t_k}), \quad t_k \in \{0, \dots, t_{N-1}\}, \tag{3c}$$

$$x_{t_0} \text{ given}, \tag{3d}$$

where the objective function is a *discounted cumulative reward* defined as:

$$J^{\pi} = r_{t_N}(x_{t_N}) + \sum_{k=0}^{N-1} \gamma^k r(x_{t_k}), \tag{4}$$

with $r : \mathcal{X} \to \mathbb{R}$ being the *reward* received by the learning agent, γ a discount factor set to 0.99, and $r_{t_N} : \mathcal{X} \to \mathbb{R}$ being the *final reward*. In particular, the reward function is formulated as a distance-like function between the bacterial density in the chamber and a given reference set point $\bar{x}$ as follows:

$$r(x_{t_k}) = -(x_{t_k} - \bar{x})^2, \tag{5}$$

which steers the learning agent towards achieving and maintaining the bacterial density at the reference setpoint value $\bar{x}$.

3. Control Design and Validation

To solve the learning problem and regulate the density of the bacterial population in a bioreactor, we designed a Deep Q-Learning algorithm leveraging the *sim-to-real* approach. Specifically, as a test-bed species, we utilized the *Escherichia coli* (*E. coli*) strain designed by Gardner et al. (2000), which embeds a plasmid implementing a genetic toggle-switch (i.e., a reversible bistable memory mechanism).

In this section, we illustrate the three-step pipeline we used to develop our control algorithm (Figure 2). First, we selected and calibrated a dynamical model capable of capturing the growth dynamics of the microorganisms. Next, this mathematical model was employed to generate synthetic data for training the neural network. Finally, the trained network was deployed *in vivo* to control the population density inside the bioreactor.

3.1. Modeling the Microbial Growth Simulator

The production of synthetic data for model training requires the selection and parameterization of a mathematical model that captures the main dynamical features of bacterial growth. An established model for describing the exponential growth of bacteria in bioreactors can be written as (Monod, 1949):

$$\dot{x}_t = \left(\mu(T) - \frac{u_t}{\tau}\right) x_t, \tag{6}$$

where x is the density of the cellular population, μ is the growth rate of the population, τ is a scaling factor, and u represents our control input (i.e. the dilution rate applied by modulating the speed of the pump carrying fresh media into the reactor). Note that the growth rate of the cells is generally

influenced by the temperature of the culture T. However, in our modeling we assumed a constant temperature, hence we fixed $\mu(T) = \mu$.

All the quantities in the above model are dimensionless. The measured optical density (OD) takes values between 0 and 1, corresponding to the absence and abundance of bacteria in the chamber, respectively, and is calibrated at the beginning of the experiments.

To parametrize this model we conducted a single open-loop experiment growing the bacteria in the Chi.Bio at different values of the dilution rate, which were changed randomly every 30 minutes. All the experiments, were performed at 37 ^{o}C in Luria broth media supplemented with $50\,\mu g/mL$ Kanamycin and $1\,mM$ Isopropyl β-D-1-thiogalactopyranoside (IPTG).

The values of μ and τ were estimated from experimental data using a least square estimator in MATLAB and validated via open-loop experiments. In these experiments, cells were grown for 60 minutes. Subsequently, the cell culture was diluted using the maximum available dilution rate of $0.02\,mL/s$ until the OD fell below 0.3. Finally, the dilution rate was randomly changed every 30 minutes.

Figure 2.a (bottom left panel) shows the trajectory generated by the model parametrized with experimental data (in orange) alongside the real data recorded from the Chi.Bio (in blue). Note that the model effectively captures both the dynamics of the exponential growth of the population and the effects of dilution. However, the prediction of the system trajectories achieved a root mean squared error (RMSE) of 7.15, which is less accurate if compared with more sophisticated models. For instance, the model employed in (Brancato et al., 2024) achieved an RMSE of 3.18. The question is now whether using such a simple model can be effective when generating synthetic data for the design of a learning-based controller to be used *in vivo*.

3.2. Training and Deployment of the Learning-Based Controller

We implement a DQN algorithm (Mnih et al., 2015) in which a neural network approximates the optimal action-value function (see Watkins and Dayan (1992)). Specifically, the neural network is used to estimate the action u based on the current OD measure x and the desired reference values $\bar{x}$ of the OD, which are the neural network inputs. The training is performed by using synthetic data generated by the simplified deterministic mathematical model (6), while $\bar{x}$ is randomly drawn with uniform distribution from the discrete set $\{0.2,\, 0.3,\, \ldots\,,\, 0.9,\, 1\}$ at each episode. By doing so, we enable the artificial agent to regulate the system on multiple set points using only a single Q-network. Similarly, for each episode, we generate initial conditions from the same discrete set as $\bar{x}$. The set of possible control actions, u, comprises 17 discrete uniformly distributed in the interval $[0, 0.02]$ representing the admissible *pump rates*. The neural network architecture includes two fully connected layers, each with 64 nodes activated by ReLU functions. We used the Adam Optimizer for training with a learning rate of 0.001. Training involved 100 *in-silico* episodes using the model (6), with each episode consisting of 100 steps corresponding to the one-minute sampling time dictated by the bioreactor's constraints. To accurately simulate the continuous-time dynamics of the cells, the synthetic OD measures, x, were generated by integrating (6) with a finer time step of 0.1 minutes. The results of the cumulative reward are depicted in Figure 2.b.

Once the DQN was trained with synthetic data, we implemented the control strategy in real-time to regulate the OD within a Chi.Bio bioreactor. This bioreactor housed a culture of *E. coli* strain embedding a plasmid with a genetic toggle switch. The time evolution of the OD, as shown in Figure 2.c, demonstrates the controller's effectiveness; the desired set point of 0.5 is successfully achieved

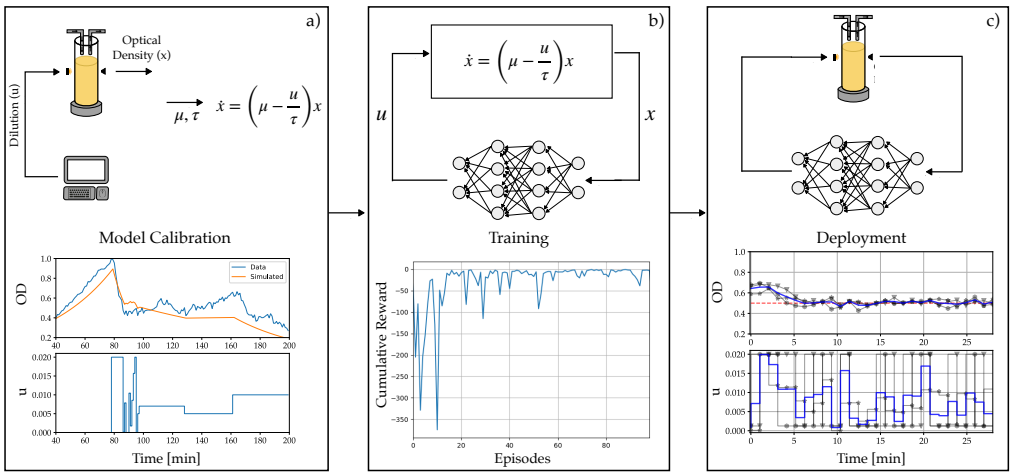

Figure 2: Sim-to-real pipeline. **Panel a)** Top: Selection and calibration of a dynamic model to describe microorganism growth. Bottom: Comparison of the calibrated model predictions (orange) and actual data collected using the Chi.Bio (blue). **Panel b)** Top: Use of the mathematical model to generate synthetic data for training the neural network. Bottom: Graph showing the progression of the cumulative reward over 100 training episodes. **Panel c)** Top: Application of the trained network to regulate the density of the cell population within the bioreactor. Bottom: Time evolution of the optical density (OD) controlled by the DQN-based algorithm, with the corresponding control inputs to the pump. Solid lines (in blue) indicate the average state and input evolution over the three *in vivo* experiments (in gray). The red dashed lines mark the set point $\bar{x}$ of 0.5.

and maintained, with an average settling time of 10 minutes. This rapid stabilization indicates the controller's efficiency in managing the system dynamics effectively.

Next, we will evaluate the performance of the proposed controller in adjusting to changes in the desired OD and its robustness against variations in the culture's temperature, which affect the intrinsic growth rate of the cells.

3.3. In Vivo Performance and Robustness Assessment

We evaluated the *sim-to-real* DQN controller's performance through multiple experiments conducted with the Chi.Bio. Following a recovery phase where the cells were allowed to grow with abundant nutrients, we diluted the culture to reach a specific target OD. This reference OD value was then maintained for 30 minutes.

The experiments were replicated three times for each desired value of OD, specifically set to 0.8, 0.65, and 0.5, respectively. The three replicates and averages of the controlled OD and the control policy, for each experimental scenario, are depicted in Figure 3.a-b 2.c.

Moreover, we assessed the robustness of the DQN controller to temperature variations, which directly affect the cells' growth rate. We observed a 10% decrease in growth rate when the temperature was reduced from the nominal condition of 37°C to 30°C. The robustness test involved starting the experiments at 37°C and regulating the OD to a target value of 0.5 for 30 minutes. Subsequently, we lowered the temperature in the Chi.Bio to 30°C. The outcomes of this experiment are displayed in Figure 3.c. Despite the perturbation in the intrinsic growth rate of the cells due to the temperature change, the controller successfully maintained the OD at the desired value, demonstrating its robustness.

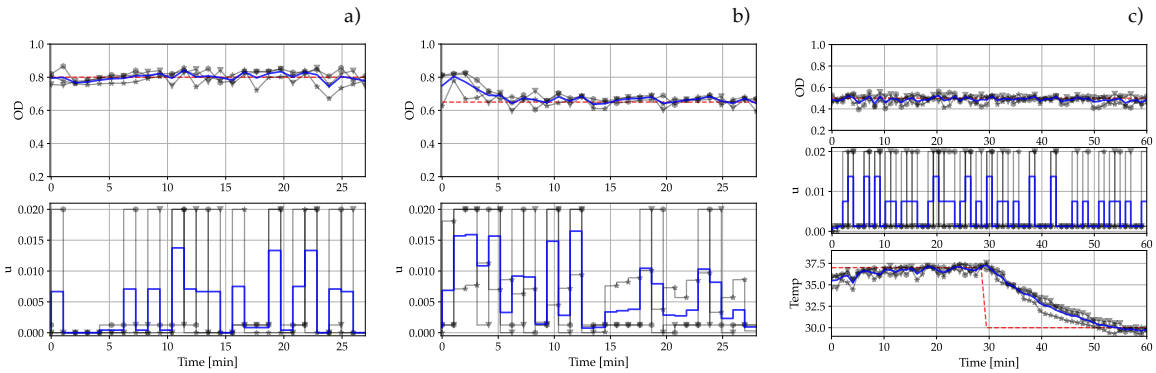

Figure 3: Performance and robustness analysis. **Panel a)** Displays the results of OD regulation for a reference of 0.8. Top subpanel: Time evolution of the OD. Bottom subpanel: Control input computed by the DQN-based controller. **Panel b)** Shows the results for a reference OD of 0.65. Top subpanel: Time evolution of the OD. Bottom subpanel: Control input computed by the DQN-based controller. **Panel c)** Demonstrates the results of OD regulation for a reference of 0.5 under a temperature change. After 30 minutes, the temperature is switched from 37°C to 30°C. Solid lines (in blue) represent the average evolution of state and input over the three *in vivo* experiments (in gray). The red dashed lines indicate the set points $\bar{x}$ of 0.8,0.65,0.5 respectively.

4. Control Benchmarks and Comparison

In what follows, we will compare the performance of our proposed learning-based controller with other common controller types used in synthetic biology applications for regulating biochemical processes. Specifically, we focus on the Proportional Integral (PI) controller and the Model Predictive Controller (MPC).

The PI controller is the one already integrated into the Chi.Bio set-up (Steel et al., 2020), while the MPC has been developed specifically for the sake of comparison. The PI strategy includes a proportional action based on the error between the desired and measured OD values. It also includes two integral actions: a traditional action designed to eliminate steady-state errors and an additional one tailored to compensate for the effects of faulty gaskets in the pumps.

The MPC determines the necessary control inputs by solving an optimization problem during each control cycle. Specifically, at each time step, the MPC solves an optimization problem over a finite prediction horizon of five minutes $T_h = 5\,\text{min}$, seeking the policy that minimizes the cost function:

$$J = \sum_{k=0}^{N-1} c_k + V_F(x_N), \tag{7}$$

where the cost term c_k is defined as:

$$c_k = \begin{cases} 100 & \text{if } u \notin [0, 0.02] \\ (x_k - \bar{x})^2 & \text{otherwise} \end{cases} \tag{8}$$

This formulation is designed to penalize both the deviation from the desired OD and any violation of the constraints on the actuators. The final cost is defined as $V_F(x_N) = (x_N - \bar{x})^2$.

The MPC uses the model described in (6) to run the optimization problem, which is resolved using a particle swarm optimizer (Bonyadi and Michalewicz, 2017). The resulting control input, calculated as the solution to this optimization problem, is then applied to the real system during the subsequent control interval of one minute.

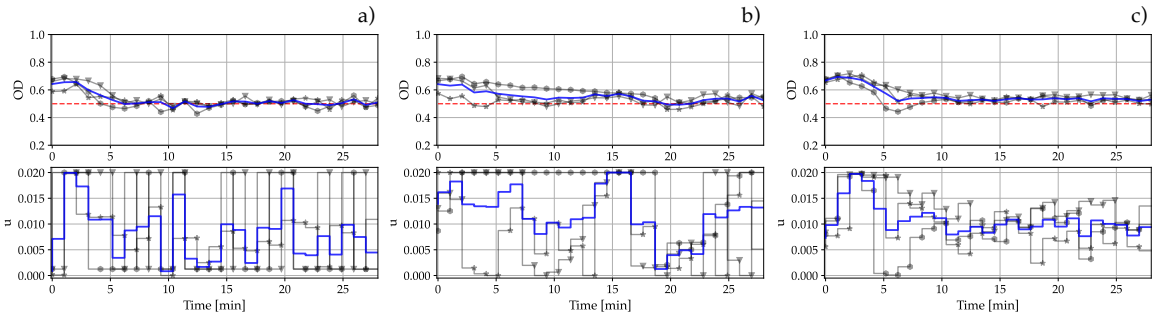

Figure 4: Controllers ' comparison: The results of the OD regulation for reference of 0.5. The top panels depict the time evolution of the OD while the bottom panels depict the control input computed by a) DQN, b) PI, c) MPC, respectively. Solid lines (in blue) represent the average evolution of state and input over the three *in vivo* experiments (in gray). The red dashed lines indicate the set point $\bar{x}$ of 0.5.

4.1. Comparison

To assess quantitatively the performance of the control algorithms we used two integral metrics, namely the Integral Squared Error (ISE) and the Integral Time Absolute Error (ITAE), that provide a quantitative measure of the transient and static performance, respectively. More precisely, the ISE and ITAE are defined as (Fiore et al., 2016; Guarino et al., 2020):

$$\text{ISE} = \frac{1}{t_f}\int_0^{t_f}(\bar{x} - x(\tau))^2 d\tau, \quad \text{ITAE} = \frac{1}{t_f}\int_0^{t_f}\tau|\bar{x} - x(\tau)|d\tau, \tag{9}$$

where $\bar{x}$ is the desired density and t_f is the duration of the experiment. The outcome of the experiments shown in Figure 4 confirms the capability of all the controllers to regulate the density of the population of interest. Furthermore, Table 1 shows the controller's performances, comparing the learning-based control strategy with the PI and the MPC, confirming the viability of a *sim-to-real* paradigm in a biological setting. Note that all the controllers have comparable performances with the DQN offering comparable performance and robustness to those of the MPC (see Table 1).

5. Discussion

In this study, we regulated the optical density of an *E. coli* population in a small turbidostat using a machine learning-based external control approach. To overcome the data efficiency issue that often renders algorithms impractical for synthetic biology applications, we adopted and experimentally validated the use of a *sim-to-real* paradigm. Specifically, the policy was initially acquired through training with a deterministic mathematical model of cell growth, which was parametrized using a limited number of experiments. Subsequently, this policy was validated through *in vivo* experimental testing, where different sources of noise are present. Our experimental results demonstrate that a learning-based control, trained on an approximate and deterministic simulator, can effectively regulate population density during *in vivo* experiments. We wish to emphasize that this work represents one of the first experimental confirmations that a sim-to-real policy can be used to regulate the density of bacterial populations. Building on the results presented here, future work will focus on enhancing the performance and robustness of the learned policy. For instance, a more accurate model could be utilized for training by explicitly modeling noise sources. Alternatively, domain

	DQN	PI	MPC
Reference 0.8			
ISE	0.039 ± 0.010	0.046 ± 0.010	**0.035** ± 0.010
ITAE	12.21 ± 1.31	12.38 ± 2.03	**10.44** ± 2.57
Reference 0.65			
ISE	**0.032** ± 0.32	0.039 ± 0.019	0.092 ± 0.061
ITAE	11.90 ± 0.91	10.50 ± 2.23	**9.38** ± 4.23
Reference 0.5			
ISE	**0.051** ± 0.034	0.111 ± 0.099	0.117 ± 0.048
ITAE	**7.43** ± 0.82	11.49 ± 4.14	12.98 ± 4.64
Temperature 37°C			
ISE	0.045 ± 0.016	0.033 ± 0.007	**0.032** ± 0.007
ITAE	12.42 ± 1.28	**9.65** ± 1.90	11.60 ± 2.09
Temperature 30°C			
ISE	0.031 ± 0.016	**0.025** ± 0.003	0.042 ± 0.015
ITAE	11.84 ± 3.64	**9.12** ± 1.29	11.54 ± 3.23

Table 1: Control performance and robustness comparison. The performance indices used are the Integral Squared Error (ISE) and the Integral Time averaged Absolute Error (ITAE). Each entry shows the average and standard deviation for each index over the $n = 3$ experimental replicates. The minimum values in each row are emphasized in boldface.

randomization on the values of μ and τ could be employed during the training phase. From a methodological perspective, we are also interested in obtaining controllers for biochemical systems from few example data, exploiting frameworks like that of Gagliardi and Russo (2022), which accounts for nonlinear, non-stationary, and stochastic dynamics. Moreover, we aim to develop and assess a sim-to-real learning-based controller that leverages differences in growth rates between two distinct cell populations to regulate their relative densities inside a bioreactor. This presents a significantly more challenging problem than those addressed with more traditional approaches, such as PI and MPC.

Acknowledgments

We express our sincere gratitude to the Telethon Institute of Genetics and Medicine (TIGEM) in Pozzuoli and Scuola Superiore Meridionale in Naples for their invaluable support and resources, which greatly contributed to the successful completion of this scientific paper. Additionally, DF and GR wish to acknowledge the financial support provided by the European Union - Next Generation EU, under PRIN 2022 PNRR, for the Project "Control of smart microbial communities for wastewater treatment."

References

Chetan Aditya, François Bertaux, Gregory Batt, and Jakob Ruess. A light tunable differentiation system for the creation and control of consortia in yeast. *Nature Communications*, 12(1):5829, 2021.

François Bertaux, Sebastián Sosa-Carrillo, Viktoriia Gross, Achille Fraisse, Chetan Aditya, Mariela Furstenheim, and Gregory Batt. Enhancing bioreactor arrays for automated measurements and reactive control with reacsight. *Nature Communications*, 13(1):3363, 2022.

Dimitri P. Bertsekas. *Dynamic Programming and Optimal Control*, volume I. Athena Scientific, Belmont, MA, USA, 3rd edition, 2005.

Mohammad Reza Bonyadi and Zbigniew Michalewicz. Particle swarm optimization for single objective continuous space problems: a review. *Evolutionary Computation*, 25(1):1–54, 2017.

Sara Maria Brancato, Davide Salzano, Davide Fiore, Giovanni Russo, and Mario di Bernardo. Ratiometric control of two microbial populations in a dual chamber bioreactor. *BioRxiv*, 2024.

Katie Brenner, Lingchong You, and Frances H Arnold. Engineering microbial consortia: a new frontier in synthetic biology. *Trends in Biotechnology*, 26(9):483–489, 2008.

Lucian Buşoniu, Tim de Bruin, Domagoj Tolić, Jens Kober, and Ivana Palunko. Reinforcement learning for control: Performance, stability, and deep approximators. *Annual Reviews in Control*, 46:8–28, 2018.

Yoojin Choi, Tae Jung Park, Doh C Lee, and Sang Yup Lee. Recombinant escherichia coli as a biofactory for various single-and multi-element nanomaterials. *Proceedings of the National Academy of Sciences*, 115(23):5944–5949, 2018.

Patrick De Leenheer and Hal Smith. Feedback control for chemostat models. *Journal of Mathematical Biology*, 46(1):48–70, 2003.

Davide Fiore, Fabio Della Rossa, Agostino Guarino, and Mario di Bernardo. Feedback ratiometric control of two microbial populations in a single chemostat. *IEEE Control Systems Letters*, 6: 800–805, 2021.

Gianfranco Fiore, Giansimone Perrino, Mario di Bernardo, and Diego di Bernardo. In vivo real-time control of gene expression: a comparative analysis of feedback control strategies in yeast. *ACS Synthetic Biology*, 5(2):154–162, 2016.

Davide Gagliardi and Giovanni Russo. On a probabilistic approach to synthesize control policies from example datasets. *Automatica*, 137:110121, 2022.

Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, 2000.

Agostino Guarino, Davide Fiore, Davide Salzano, and Mario di Bernardo. Balancing cell populations endowed with a synthetic toggle switch via adaptive pulsatile feedback control. *ACS Synthetic Biology*, 9(4):793–803, 2020.

Joaquín Gutiérrez Mena, Sant Kumar, and Mustafa Khammash. Dynamic cybergenetic control of bacterial co-culture composition via optogenetic feedback. *Nature Communications*, 13(1):4808, 2022.

Roland Hafner and Martin Riedmiller. Reinforcement learning in feedback control: Challenges and benchmarks from technical process control. *Machine learning*, 84:137–169, 2011.

Joachim J Hug, Daniel Krug, and Rolf Müller. Bacteria as genetically programmable producers of bioactive natural products. *Nature Reviews Chemistry*, 4(4):172–193, 2020.

Stephen James, Andrew J Davison, and Edward Johns. Transferring end-to-end visuomotor control from simulation to real world for a multi-stage task. In *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pages 334–343. PMLR, 2017.

David Jullesson, Florian David, Brian Pfleger, and Jens Nielsen. Impact of synthetic biology and metabolic engineering on industrial production of fine chemicals. *Biotechnology Advances*, 33 (7):1395–1402, 2015.

Minori Kusuda, Hiroshi Shimizu, and Yoshihiro Toya. Reactor control system in bacterial co-culture based on fluorescent proteins using an arduino-based home-made device. *Biotechnology Journal*, 16(12):2100169, 2021.

Jean-Baptiste Lugagne, Caroline M Blassick, and Mary J Dunlop. Deep model predictive control of gene expression in thousands of single cells. *Nature Communications*, 15(1):2148, 2024.

Yongkun Lv, Shuai Qian, Guocheng Du, Jian Chen, Jingwen Zhou, and Peng Xu. Coupling feedback genetic circuits with growth phenotype for dynamic population control and intelligent bioproduction. *Metabolic Engineering*, 54:109–116, 2019.

Marco Mauri, Jean-Luc Gouzé, Hidde De Jong, and Eugenio Cinquemani. Enhanced production of heterologous proteins by a synthetic microbial community: Conditions and trade-offs. *PLOS Computational Biology*, 16(4):e1007795, 2020.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

Jacques Monod. The growth of bacterial cultures. *Annual review of microbiology*, 3(1):371–394, 1949.

Benjamin Recht. A tour of reinforcement learning: The view from continuous control. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:253–279, 2019.

Andrei A. Rusu, Matej Večerík, Thomas Rothörl, Nicolas Heess, Razvan Pascanu, and Raia Hadsell. Sim-to-real robot learning from pixels with progressive nets. In *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pages 262–270. PMLR, 2017.

Tulasi Satyanarayana. *Yeast Biotechnology: Diversity and Applications*. Springer, 2009.

Harrison Steel, Robert Habgood, Ciarán L Kelly, and Antonis Papachristodoulou. In situ characterisation and manipulation of biological systems with chi. bio. *PLOS Biology*, 18(7):e3000794, 2020.

Yuan Su, Chuan Liu, Huan Fang, and Dawei Zhang. Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microbial cell factories*, 19(1):1–12, 2020.

Jie Tan, Tingnan Zhang, Erwin Coumans, Atil Iscen, Yunfei Bai, Danijar Hafner, Steven Bohez, and Vincent Vanhoucke. Sim-to-real: Learning agile locomotion for quadruped robots. *arXiv preprint arXiv:1804.10332*, 2018.

Rongzhen Tian, Yanfeng Liu, Yanting Cao, Zhongjie Zhang, Jianghua Li, Long Liu, Guocheng Du, and Jian Chen. Titrating bacterial growth and chemical biosynthesis for efficient n-acetylglucosamine and n-acetylneuraminic acid bioproduction. *Nature Communications*, 11(1): 5078, 2020.

Neythen J Treloar, Alex JH Fedorec, Brian Ingalls, and Chris P Barnes. Deep reinforcement learning for the control of microbial co-cultures in bioreactors. *PLOS Computational Biology*, 16(4): e1007783, 2020.

Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3):279–292, 1992.

Brandon G Wong, Christopher P Mancuso, Szilvia Kiriakov, Caleb J Bashor, and Ahmad S Khalil. Precise, automated control of conditions for high-throughput growth of yeast and bacteria with evolver. *Nature Biotechnology*, 36(7):614–623, 2018.

Haiyang Xu, Daniel Lybrand, Stefan Bennewitz, Alain Tissier, Robert L Last, and Eran Pichersky. Production of trans-chrysanthemic acid, the monoterpene acid moiety of natural pyrethrin insecticides, in tomato fruit. *Metabolic Engineering*, 47:271–278, 2018.

Ruoqi Zhang, Per Mattsson, and Torbjörn Wigren. Robust nonlinear set-point control with reinforcement learning. In *2023 American Control Conference (ACC)*, pages 84–91. IEEE, 2023.

Jianguo Zhao, Chunyu Yang, Weinan Gao, and Linna Zhou. Reinforcement learning and optimal setpoint tracking control of linear systems with external disturbances. *IEEE Transactions on Industrial Informatics*, 18(11):7770–7779, 2022.

Guang-Yan Zhu, Abdelqader Zamamiri, Michael A Henson, and Martin A Hjortsø. Model predictive control of continuous yeast bioreactors using cell population balance models. *Chemical Engineering Science*, 55(24):6155–6167, 2000.