Applying Reinforcement Learning to Dissolved Oxygen Control in a Fed-Batch Fermentation using a Digital Twin

Rasmus Høst Villadsen[1], Krist V. Gernaey[1], Beichen Zhao[2], Wei Yu[2] and Brent Young[2]

***Abstract*— Digital twins can enable closed-loop optimisation in biomanufacturing, but reinforcement learning (RL) controllers are highly sensitive to the training environment used in simulation. In bioprocess modeling, purely data-driven environments may achieve low prediction error while violating physical constraints, which can degrade control performance. This paper investigates how environment choice affects RL for dissolved oxygen (DO) control in an *Escherichia coli* fed-batch fermentation. We develop (i) a data-calibrated mechanistic model with Monod-type kinetics and time-varying oxygen transfer, and (ii) a hybrid digital model in which mechanistic features are used as inputs to an LSTM that predicts DO one step ahead. We compare model quality and robustness on unseen data and then train RL agents within each environment to stabilise DO under pulse feeding. The study highlights the trade-off between predictive accuracy and physical consistency when using hybrid models for RL training and provides guidance for selecting simulation environments that yield robust control policies.**

I. INTRODUCTION

Sustainability goals and the rising demand for bio-based products are driving rapid progress in biomanufacturing. A key step toward modern biomanufacturing was the work of Cohen and Boyer in 1973. They constructed recombinant plasmids carrying foreign genes in *Escherichia coli* [1]. This enabled microbial production of industrial enzymes and therapeutic proteins, including human insulin [2], [3]. More recently, synthetic biology has enabled the design of improved and even novel biosynthetic pathways. This shift has expanded the applications in biofuels, food ingredients, and biopharmaceuticals, and it continues to drive growth in the biosolutions sector [4]–[7].

Industrial bioprocess control is complex. To achieve high product yield and productivity, fermentations must be monitored and controlled near their best operating conditions [8], [9]. In practice, the bioprocess industry often remains below this level of digital maturity. A main reason is the stochastic nature of biological systems and the limited understanding of intracellular metabolism [10]. Data-grounded dynamic models (digital twins) have been used at full scale to reproduce fast DO and nitrogen-species dynamics under varying aeration patterns, enabling scenario analysis and mitigation strategies [11]. However, metabolic pathways can be highly sensitive to changes in operating conditions [12], [13], this makes accurate kinetic modeling difficult. One way to improve prediction is to combine mechanistic models with data-driven models [14]–[16]. Such hybrid models can use plant data to improve simulation accuracy (often described as a digital shadow) [14], [17]. Better prediction of process states supports better control decisions and reduces plant–model mismatch [18].

Reinforcement learning (RL) has also gained attention in bioprocess research. RL can handle nonlinear and stochastic control problems, and it has been used for parameter determination [19] and process control [12], [20]–[24]. For example, Petsagkourakis et al. used a policy-gradient method for bioprocess optimisation in a simulated setting with disturbances [12]. However, their state variables included substrate, product, and biomass, which are usually not available online. Espinel-Ríos et al. studied RL architecture improvements using *E. coli* simulation case studies and a policy-gradient method [24]. Treloar et al. applied neural fitted Q-learning to optimise bang–bang feeding for an *E. coli* co-culture, using a Monod-type model as the environment [20]. For RL, these the training environments matter because the it shapes what the agent learns. If the environment is not robust, the learned policy may fail when applied to unseen runs or disturbances.

Most RL studies in bioprocess control train agents using *in silico* data from mechanistic models with simplified dynamics or arbitrarily chosen parameters. Many studies also assume that key state variables, such as biomass, substrate, and product concentrations, can be measured at any time. These assumptions limit implementation and validation for two reasons. First, simplified mechanistic simulations may not represent real processes well. They can miss real disturbances, nonlinear behavior, and changing dynamics. As a result, learned policies can be fragile and hard to transfer to plant operation. Second, the RL agent must observe the system states to choose actions. Direct online measurement of these state variables is rarely available in practice [14], [25], [26].

To address this issue, RL training environments can be built from digital twins calibrated with real fermentation data. One option is a data-grounded mechanistic model, which helps the environment reflect real process behaviour and preserves physical structure. Another option is a hybrid approach that combines a mechanistic backbone with a data-driven model trained on process data, which can capture complex dynamics that are difficult to represent purely mechanistically.

In this work, we develop two RL training environments from real fed-batch fermentation data. The first is a data-grounded mechanistic model fitted to biomass, substrate, and dissolved oxygen (DO) measurements, providing an interpretable description of the dominant process dynamics. The

[1]Department of Chemical and Biochemical Engineering, Technical University of Denmark, Denmark

[2]Department of Chemical and Materials, University of Auckland, New Zealand

second is a hybrid model in which a long short-term memory (LSTM) network predicts DO dynamics from mechanistic features (e.g., biomass/substrate rates and aeration variables).

We then test how the learned policies depend on the training environment by evaluating each trained controller in closed-loop simulations and benchmarking DO stability against a baseline strategy. This isolates whether better model fit leads to better control, and whether the policy generalises beyond environment-specific artefacts.

II. Methods

A. Fermentation data

Fermentation data are obtained from experiments using *E. coli* cultivated in fed-batch mode in a bang-bang feeding control. A concentrated glucose solution (500 g/L) is fed at 0.5 mL/min whenever DO (in % saturation) rises to 60%, and feeding is stopped when DO falls to 25%. The DO in %sat is converted to concentration (c_O^*) by the following formula [27]:

$$c_O^* = 0.027 \cdot e^{\frac{1142}{T}} \cdot \pi \tag{1}$$

where the temperature T is 310.15 K and π is the partial pressure of oxygen (0.2095 atm). This gives a saturation concentration at 0.225 mM at 37 °C. The initial conditions of the fermentation process are listed in Table I.

TABLE I
Fermentation initial conditions

Parameter	Description	Value	Unit
X_0	Biomass concentration	0.12	g/L
S_0	Substrate concentration	10.66	g/L
P_0	Product concentration	0	g/L
DO_0	Oxygen tension	0.225×0.88	mmol/L
V_0	Working volume	0.70	L

B. Batch fermentation modeling

1) Mechanistic model: A dynamic model of a fermentation can be used to forecast the evolution of state variables of the fermentation, acting as a digital shadow [28]. The substrate and biomass can be modeled to follow Monod-type kinetics with oxygen dependency, given by [29]:

$$\frac{dX}{dt} = \mu_{max} \cdot \frac{S}{S + K_S} \cdot \frac{c_O}{c_O + K_O} \cdot X \tag{2}$$

where dX/dt is the cell growth rate, μ_{max} is the maximum specific growth rate, S is the substrate concentration, K_S is the Monod half-saturation constant, c_O the concentration of dissolved oxygen, K_O the Monod half-saturation constant for oxygen, and X is the biomass concentration. For the substrate,

$$\frac{dS}{dt} = -q_s \cdot X \tag{3}$$

where q_s is given by:

$$q_s = \left(q_{s,max} \cdot \frac{S}{S + K_S} \cdot \frac{c_O}{c_O + K_O} \right) \tag{4}$$

with $q_{s,max}$ being the maximum specific substrate consumption rate. Lastly, for the product as well,

$$\frac{dP}{dt} = \frac{Y_{PS}}{Y_{XS}} \cdot \frac{dX}{dt} \tag{5}$$

where Y_{PS} product yield on substrate and Y_{XS} is the biomass yield on substrate. Equations 2-5 describe the dynamics of a batch process, while fed-batch models take dilution into account [30], [31]. However, in fed-batch processes where feeding is given in a on/off form by bolus injections (like bang-bang control), it is not possible to model the dilution continuously. The dilution effect is instead taken into account after discretising the ODEs (Eq. 6, 2, and 3) and modeling the addition of a substrate shot. In the current study, the feeding logic applied is based on DO. That is, if DO reaches 60 %sat, substrate shots are added until the DO falls below 25 %sat again, at which point the controller waits. This logic ensures that the population is replenished with substrate when depleted (indicated by a high DO). The simulated substrate concentration is updated discretely at each time step by adding a substrate shot when the feeding condition is met. Since this feeding logic consistently updates the substrate based on online measurements, the biomass is also dependent on the data measurements and is simulated likewise. Modeling the DO can be done mechanistically as suggested by literature [32]–[34]:

$$\frac{dc_O}{dt} = k_L a \cdot (c_O^* - c_O) - q_o \cdot X \tag{6}$$

where $k_L a$ is the oxygen mass transfer coefficient and q_o is the oxygen-dependent substrate consumption rate, calculated as

$$q_o = (q_s + q_m) \cdot Y_{OS} \tag{7}$$

with q_m being the maintenance coefficient ($g/(g \cdot h)$), and Y_{OS} is the oxygen yield per unit substrate.

Since process conditions (for example aeration, agitation and broth viscocity) may vary over time, $k_L a$ is time-dependent during a fed-batch fermentation. However, if $k_L a$ is estimated, the oxygen can be simulated and respond to changes in substrate and biomass dynamics (Fig. 1).

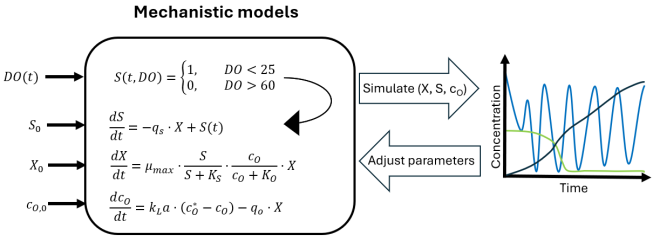

Fig. 1. The mechanistic modeling framework.

2) Hybrid model: Biomass, substrate, and product dynamics are relatively well captured by mechanistic models, but DO is more difficult to model due to time-varying mass transfer, metabolic shifts, and sensor delays. To better represent these effects, a hybrid model is developed that combines mechanistic kinetics with a data-driven grey-box

term based on a long short-term memory (LSTM) network. LSTMs are recurrent neural networks with memory that capture temporal dependencies and predict the next state from past information [35]. In this work, the mechanistic model provides physically meaningful features, while the LSTM learns the remaining DO dynamics from data [36], [37] (Fig. 2).

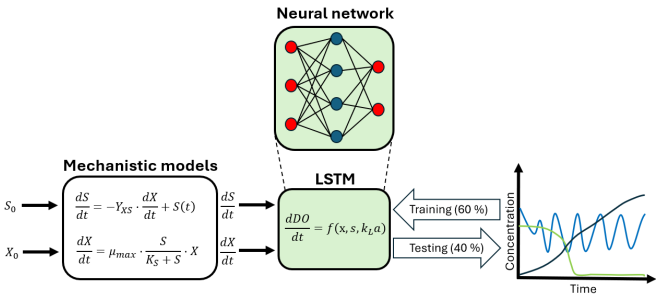

Fig. 2. Hybrid model structure for the fermentation process.

At each time step, biomass, substrate, and product are simulated using the Monod-based equations (Eqs. 2–5). From the current state (X_t, S_t) the mechanistic rates $(dX/dt, dS/dt)$ are computed and form the feature vector

$$\mathbf{u}_t = [\, dX/dt,\ X_t,\ dS/dt,\ k_L a(t)\,].$$

This feature vector is then used by the LSTM to predict the DO at the next time step, DO_{t+1}. Using current mechanistic features to predict the next DO preserves temporal causality and avoids circular dependencies.

All input features are stacked and normalised before training. The dataset is split into training and testing sets (60/40) and arranged into sequences of 50 samples (look-back window). The LSTM includes one hidden layer with 16 units and a single output node for DO. The network is trained for 100 epochs with a batch size of 64. The model is implemented in PyTorch, and the fitted scalers are saved for later use.

C. Reinforcement learning control

Control problem design The control problem is formulated as a discrete Markov decision process consisting of the state space, action space, transition probability, a reward function, and finally the discount factor (S, A, P, R, γ). S consists of six dimensions of process variables (biomass, substrate, product, oxygen, volume, time). A is discretely defined as $A = \{\textbf{wait},\ \textbf{1 mL},\ \textbf{1.5 mL}\}$ dosing. P is determined by the mechanistic/hybrid model. A multi-objective reward function (R) is designed that rewards stable DO but penalises excessive and insufficient substrate feeding as this would also result in stable DO. The final component of the MPD is the discount factor γ, which is set to 0.95 to balance immediate and future rewards.

Learning algorithm and architecture Deep Q-Networks (DQN) is used as a learning algorithm for the agent due to its simple application and discrete action spaces. The neural network consists of an input layer of 6 nodes (state dimension), hidden layer of 4 nodes (relatively low complexity), and finally an output layer of 3 nodes determined by the action space. ϵ-greedy exploration is balanced by applying an initial $\epsilon = 1$ with a decay rate of 0.995 per episode to exploit rewarding actions in later episodes. Experience replay (the memory of the agent) is managed by having a buffer size of 2000 transitions (default), and a batch size of 32 samples (default). A learning rate of $\alpha = 0.001$ is used (default). Training is performed over 500 episodes, with each episode representing a complete fermentation run. However, the fermentation run is cut to the initial 14 hours to avoid excessive computational expense of the remaining process, which is regarded as repetitions. Training progress is evaluated as the total rewards of one episode (the return). The agent's training is then evaluated by its policy performance.

Benchmarking environments and controllers The fidelity of the models is assessed through their ability to capture the dynamics between substrate feeding and DO response. When substrate is introduced to an actively growing culture, the immediate consequence is an increase in oxygen uptake rate, showing a characteristic decline in DO. This fundamental physiological response serves as a critical benchmark for evaluating whether the models accurately reproduce the underlying biological system.

Controller performance is evaluated by examining oxygen process trajectories under closed-loop control. The number of oscillation cycles and the amplitude of DO are used as a performance indicator. Peaks and lows are identified within this region using the find_peaks function from the SciPy library. Then, the high-to-low amplitude is calculated for each oscillation cycle and is averaged to constitute the performance indicator together with the number of oscillation cycles within that period. Here, 6-14 hours are chosen, since the controller is active in this window, and to reduce the computational expense of simulating the remaining process.

III. Results and Discussion

A. Mechanistic modeling

Modeling the sharp rises and drops of the DO required careful fine tuning of parameters. The workflow depicted in Fig. 1 is used to model the biomass and substrate trajectories (Fig. 3A-B), and the time-dependent $k_L a$ is calculated using Van't Riet's equation [38] to simulate the oxygen profile (Fig. 3C-D).

The time-dependent $k_L a$ was used to predict the sudden rises in the DO at 2.4 hours and 4.4 hours which resulted from an increase in agitation and aeration rate, and is considered to be fairly accurate (RMSE = 0.081). It was observed that the overall trend of the simulated DO showed greater oscillation amplitudes than the real data (Fig. 3D), possibly due to change in viscocity not reflected in the model. However, the simulation converged well the to measured DO regarding the instantaneous jumps due to increased $k_L a$ in the initial hours and the sharp oscillations during pulse feeding in the later hours. The simulated biomass and

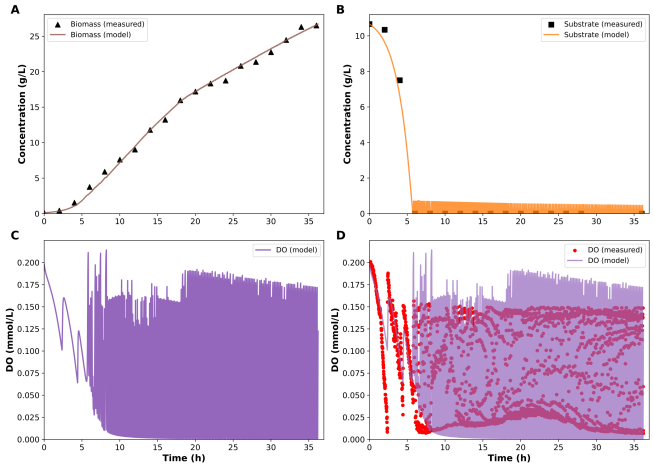

Fig. 3. Mechanistic model predictions: A) Biomass, B) Substrate, C-D) Oxygen profiles, using parameters that resulted in a low RMSE between modeled and measured dissolved oxygen.

substrate profile also matched the measured data well (Fig. 3A-B), although the substrate was undetectable after 5 hours.

B. Hybrid modeling

Pure mechanistic modeling of such complex oxygen dynamics is nontrivial and with the risk of overfitting parameters to measurement noise. For these reasons, a data-driven hybrid approach is constructed using LSTM to predict the DO. Here, a set of parameters and models for biomass and substrate are used as input features to the model, that uses these to predict the DO (Fig. 4D). The dashed line indicates the split in training and testing predictions for the DO. The first 60 % of the DO data were used for DO training predictions and the remaining 40 % for DO testing predictions.

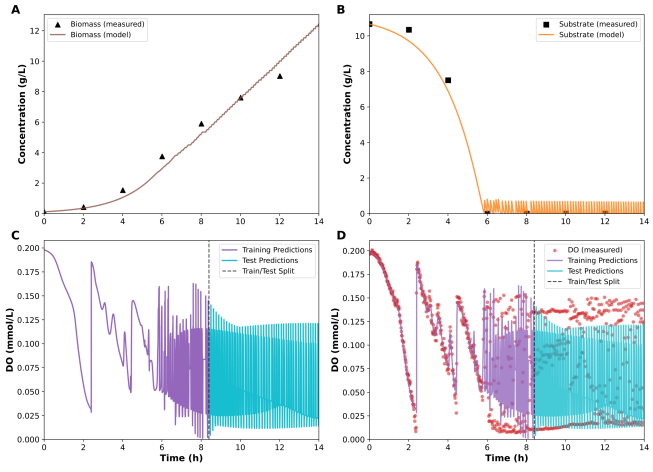

Fig. 4. Hybrid model predictions: A) Biomass, B) Substrate, C-D) Dissolved oxygen (DO). The dashed line indicates the split of training and testing data.

Looking at Fig. 4D, the training predictions from 0-6 hours followed the measured DO very accurately, whereas the remaining training and model evaluation (testing predictions) were less accurate. However, the model still appeared to capture the oscillating trend in the oxygen profile well (test RMSE = 0.051). Comparing with the mechanistic model, the hybrid performed better when considering RMSE solely. When determining which model is better as an RL environment, other factors must be considered. For example, the delay from the DO sensor signal to pump activation to increased metabolic activity is not represented in the simulation as in the real data. Thus, a small offset may have a large impact on the RMSE in the mechanistic approach.

C. Control using Reinforcement learning

1) Training: The training progress of the agent trained on the mechanistic environment is illustrated in Fig. (5). The

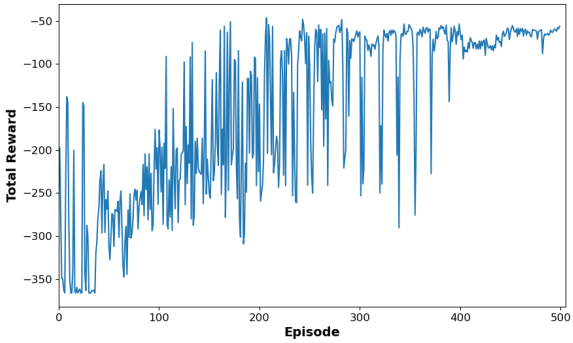

Fig. 5. Training progress of the RL agent based on the mechanistic model using 500 episodes. One episode corresponds to a fermentation simulation from 0-14 hours.

training progress indicates that the agent picks up an optimal policy within 500 episodes, since the return is not increasing hereafter. Therefore, 500 episodes were the training load for further runs to reduce computational expense. The return has a large variance at the beginning of the training, indicating an unstable training. However, the return is reaching a stable plateau, which can be explained as the trade-off between exploration and exploitation.

2) RL perfomance based on the mechanistic model: Figure 6 shows the decision making of the reinforcement learning agent from the last episode of training using the mechanistic model as an environment. The rapid jumps in the DO at 2.4 hours and 4.4 hours are caused by increased $k_L a$. This control strategy results in 6 oscillations after 6 hours and an average amplitude of 0.167 mmol/L. The performance of the reinforcement learning-based controller was compared with a standard bang-bang fed-batch fermentation. Compared to the bang-bang controlled process, the RL-controlled simulated fermentation demonstrated a more robust control, exhibiting less oscillation cycles, reduced by 93 %.

3) RL performance based on the Hybrid model: Figure 7 shows the decision making results of the reinforcement learning agent based on the hybrid model as an environment. Looking at the DO process trajectories (Fig. 7), the decision making of the RL agent appears more obscure and it is difficult to elucidate a systematic pattern between DO and

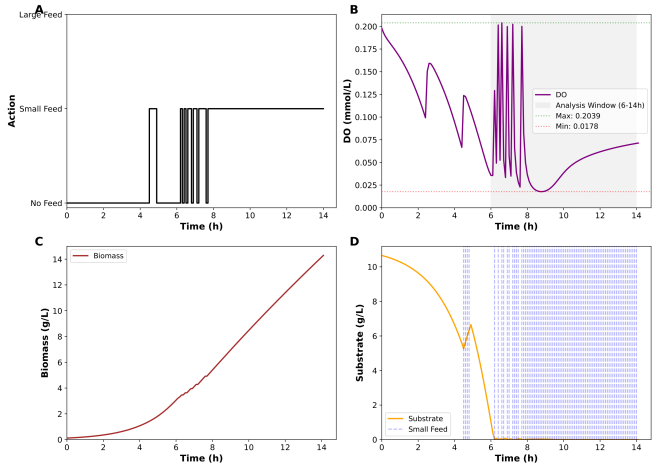

Fig. 6. RL control performance based on the mechanistic model: A) Actions taken, B) DO, C) Biomass and D) Substrate profile (the small feed option corresponds to 1 mL, and the large dosage is 1.5 mL).

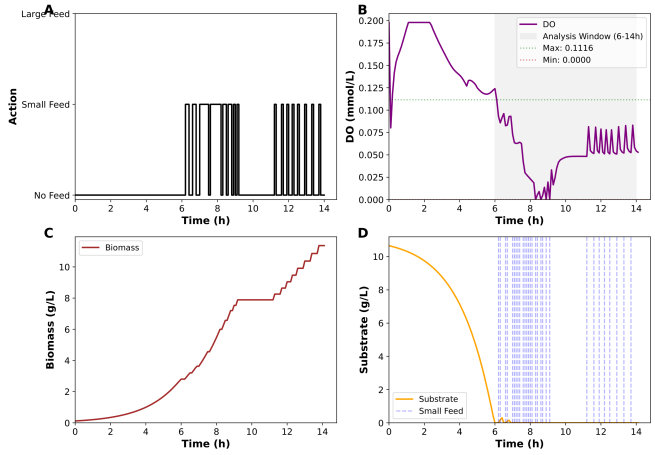

Fig. 7. RL performance: A) Actions taken (The small feed option corresponds to 1 mL, and the large dosage is 1.5 mL), B) Dissolved oxygen, C) Biomass, D) Substrate profile.

substrate availability, as was the case for the mechanistic environment. Still, the DO was accurately jumping from a low level (9.5 hours) to a higher level as there was no longer substrate consumption. Besides, there was no physically plausible relationship between the DO and substrate concentration, and the mild oscillations towards the end appeared logically reversed. That is, a small substrate addition caused a small spike in the DO. This finding indicates that the LSTM hybrid critically failed to capture true dynamics between oxygen and substrate concentrations. Small temporal offsets between feeding, oxygen uptake, and measured DO may therefore be learned as correlations rather than causal process dynamics. Hence, the hybrid could potentially perform better if it was trained with richer features, more training data and if the oxygen data were less delayed and less oscillating.

Considering the mechanistically-derived environment, the initial oxygen increase at hours 2.4 and 4.4 was well reflected by the increase in $k_L a$, and the low oxygen tension was the result of constant co-consumption of oxygen and substrate. The LSTM-derived environment, on the other hand, was less physically correct in this context (Fig. 7). Hence, while the LSTM-based hybrid model achieved a better fit on the fermentation data, it risks learning data artifacts and sensor delays instead of true oxygen-substrate dynamics. This supports the broader view that model assessment must be aligned with intended use (e.g., decision-making and control), rather than relying only on short-horizon predictive accuracy [39].

IV. CONCLUSIONS

This paper investigated reinforcement learning (RL) control of dissolved oxygen (DO) in an E. coli fed-batch fermentation using two data-grounded digital twin environments. A calibrated mechanistic model and a hybrid LSTM-based model were developed from the same experimental dataset. Although the hybrid model achieved lower DO prediction error, it did not preserve the expected physical relationship between substrate feeding and DO response.

The DQN controller trained in the mechanistic environment produced more robust DO control. It reduced DO oscillation cycles by 93% compared with the baseline bang-bang feeding strategy and also reduced oscillation amplitude. In contrast, the controller trained in the hybrid environment performed poorly because the learned DO dynamics were less physically consistent.

These results show that prediction accuracy alone is not sufficient when selecting RL training environments for bioprocess control. The environment must also preserve meaningful action–state relationships. Future work will include physically constrained hybrid models, sensing and actuation delays, parameter uncertainty, comparison with PID and MPC, and validation in real fermentation experiments.

ACKNOWLEDGMENTS

This research has been supported by the programme “Ahuora: Delivering sustainable industry through smart process heat decarbonisation”, an Advanced Energy Technology Platform, funded by the New Zealand Ministry of Business, Innovation and Employment. The author also gratefully acknowledges financial support from the foundations of fam. Lunding Uddannelsesfond, fam. Haynmans, Knud Højgaards, William Demant, Olsens Mindefond, and Otto Mønsteds, which supported travel and accommodation during the research stay in New Zealand.

REFERENCES

[1] S. N. Cohen, A. C. Y. Chang, H. W. Boyer, and R. B. Helling, “Construction of biologically functional bacterial plasmids in-vitro,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 70, no. 11, pp. 3240–3244, 1973.

[2] S. Jorgensen, K. W. Skov, and B. Diderichsen, “Cloning, sequence, and expression of a lipase gene from pseudomonas-cepacia - lipase production in heterologous hosts requires 2 pseudomonas genes,” *Journal of Bacteriology*, vol. 173, no. 2, pp. 559–567, 1991.

[3] L. Thim, M. Hansen, K. Norris, I. Hoegh, E. Boel, J. Forstrom, G. Ammerer, and N. Fiil, “Secretion and processing of insulin precursors in yeast,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 83, no. 18, pp. 6766–6770, 1986.

[4] K. Pandey, M. Pandey, V. Kumar, U. Aggarwal, and B. Singhal, "Bioprocessing 4.0 in biomanufacturing: paving the way for sustainable bioeconomy," *Systems Microbiology and Biomanufacturing*, vol. 4, no. 2, pp. 407–424, 2024.

[5] C. A. Voigt, "Synthetic biology 2020–2030: six commercially-available products that are changing our world," *Nature Communications*, vol. 11, no. 1, p. 6379, 2020.

[6] A. S. Khalil and J. J. Collins, "Synthetic biology: Applications come of age," *Nature Reviews Genetics*, vol. 11, no. 5, pp. 367–379, 2010.

[7] D. E. Cameron, C. J. Bashor, and J. J. Collins, "A brief history of synthetic biology," *Nature Reviews Microbiology*, vol. 12, no. 5, pp. 381–390, 2014.

[8] O. Konzock and J. Nielsen, "Trying to evaluate production costs in microbial biotechnology," *Trends in Biotechnology*, vol. 42, no. 11, pp. 1339–1347, 2024.

[9] C. L. Gargalo, P. C. Lopez, A. Hasanzadeh, I. A. Udugama, and K. V. Gernaey, "On-line monitoring of process parameters during fermentation," *Current Developments in Biotechnology and Bioengineering*, pp. 117–164, 2022.

[10] H. Narayanan, M. F. Luna, M. von Stosch, M. N. Cruz Bournazou, G. Polotti, M. Morbidelli, A. Butté, and M. Sokolov, "Bioprocessing in the digital age: The role of process models," *Biotechnology Journal*, vol. 15, no. 1, p. 1900172, 2020.

[11] T. Lei, J. Whale-Obrero, S. B. Larsen, K. Kjellberg, K. V. Gernaey, and X. Flores-Alsina, "Dynamically predicting nitrous oxide emissions in a full-scale industrial activated sludge reactor under multiple aeration patterns and cod/n ratios," *Water Research*, vol. 278, p. 123379, 2025.

[12] P. Petsagkourakis, I. O. Sandoval, E. Bradford, D. Zhang, and E. A. del Rio-Chanona, "Reinforcement learning for batch bioprocess optimization," *Computers and Chemical Engineering*, vol. 133, p. 106649, 2020.

[13] J. Thierie, "Modeling threshold phenomena, metabolic pathways switches and signals in chemostat-cultivated cells: The crabtree effect in saccharomyces cerevisiae," *Journal of Theoretical Biology*, vol. 226, no. 4, pp. 483–501, 2004.

[14] P. C. Lopez, I. A. Udugama, S. T. Thomsen, C. Roslander, H. Junicke, M. Mauricio-Iglesias, and K. V. Gernaey, "Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation," *Biofuels, Bioproducts and Biorefining*, vol. 14, no. 5, pp. 1046–1060, 2020.

[15] L. Riezzo, H. Kay, Y. Feng, K. Jing, and D. Zhang, "Accelerating bioprocess digital twin development by integrating hybrid modelling with transfer learning," *Chemical Engineering Journal*, vol. 511, p. 162018, 2025.

[16] D. Solle, B. Hitzmann, C. Herwig, M. Pereira Remelhe, S. Ulonska, L. Wuerth, A. Prata, and T. Steckenreiter, "Between the poles of data-driven and mechanistic modeling for process operation," *Chemie-ingenieur-technik*, vol. 89, no. 5, pp. 542–561, 2017.

[17] K. U. Rehman, X. L. Zhu, B. Wang, M. Shahzad, H. Ahmad, M. Abubakar, and M. Ajmal, "Soft sensor model based on iba-lssvm for photosynthetic bacteria fermentation process," *International Journal Bioautomation*, vol. 25, no. 2, pp. 145–158, 2021.

[18] I. A. Udugama, P. C. Lopez, C. L. Gargalo, X. Li, C. Bayer, and K. V. Gernaey, "Digital twin in biomanufacturing: challenges and opportunities towards its implementation," *Systems Microbiology and Biomanufacturing*, vol. 1, no. 3, pp. 257–274, 2021.

[19] M. R. Mowbray, C. Wu, A. W. Rogers, E. A. Rio-Chanona, and D. Zhang, "A reinforcement learning-based hybrid modeling framework for bioprocess kinetics identification," *Biotechnology and Bioengineering*, vol. 120, no. 1, pp. 154–168, 2023.

[20] N. J. Treloar, A. J. Fedorec, B. Ingalls, and C. P. Barnes, "Deep reinforcement learning for the control of microbial co-cultures in bioreactors," *Plos Computational Biology*, vol. 16, no. 4, p. e1007783, 2020.

[21] G. Bujgoi, D. Sendrescu, and D. Gagiu, "Reinforcement learning for the control of bacterial growth bioprocess," *2021 25th International Conference on System Theory, Control and Computing, Icstcc 2021 - Proceedings*, pp. 525–529, 2021.

[22] J. W. Kim, B. J. Park, T. H. Oh, and J. M. Lee, "Model-based reinforcement learning and predictive control for two-stage optimal control of fed-batch bioreactor," *Computers and Chemical Engineering*, vol. 154, p. 107465, 2021.

[23] M. Monteiro and C. Kontoravdi, "Bioprocess control: A shift in methodology towards reinforcement learning," *Computer Aided Chemical Engineering*, vol. 53, pp. 2851–2856, 2024.

[24] S. Espinel-Rios, J. L. Avalos, E. A. D. R. Chanona, and D. Zhang, "Reinforcement learning for efficient and robust multi-setpoint and multi-trajectory tracking in bioprocesses," *Computers and Chemical Engineering*, 2025.

[25] P. Cabaneros Lopez, H. Feldman, M. Mauricio-Iglesias, H. Junicke, J. K. Huusom, and K. V. Gernaey, "Benchmarking real-time monitoring strategies for ethanol production from lignocellulosic biomass," *Biomass and Bioenergy*, vol. 127, p. 105296, 2019.

[26] L. Mears, S. M. Stocks, G. Sin, and K. Gernaey, "A review of control strategies for manipulating the feed rate in fed-batch fermentation processes," *Journal of Biotechnology*, vol. 245, pp. 34–46, 2017.

[27] J. Villadsen, J. Nielsen, and G. Liden, *Bioreaction Engineering Principles, Third Edition*. SPRINGER, 2011.

[28] C. L. Gargalo, I. Udugama, K. Pontius, P. C. Lopez, R. F. Nielsen, A. Hasanzadeh, S. S. Mansouri, C. Bayer, H. Junicke, and K. V. Gernaey, "Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes," *Journal of Industrial Microbiology and Biotechnology*, vol. 47, no. 11, pp. 947–964, 2020.

[29] H. S. Fogler, *Elements of chemical reaction engineering*. Prentice-Hall, Pearson, 2016.

[30] E. Anane, A. Sawatzki, P. Neubauer, and M. N. Cruz-Bournazou, "Modelling concentration gradients in fed-batch cultivations of e. coli – towards the flexible design of scale-down experiments," *Journal of Chemical Technology and Biotechnology*, vol. 94, no. 2, pp. 516–526, 2019.

[31] C. Martínez and J. L. Gouzé, "Optimal control of a fed-batch reactor with overflow metabolism," *Ifac-papersonline*, vol. 53, no. 2, pp. 16820–16825, 2020.

[32] P. Pably, J. K. Huusom, and J. Kager, "Applying an mpc algorithm to the dissolved oxygen level of an intermittent fed-batch process," *Research Portal Denmark*, 2024.

[33] K. G. Gadkar, S. Mehra, and J. Gomes, "On-line adaptation of neural networks for bioprocess control," *Computers and Chemical Engineering*, vol. 29, no. 5, pp. 1047–1057, 2005.

[34] J. W. Kim, N. Krausch, J. Aizpuru, T. Barz, S. Lucia, P. Neubauer, and M. N. Cruz Bournazou, "Model predictive control and moving horizon estimation for adaptive optimal bolus feeding in high-throughput cultivation of e. coli," *Computers and Chemical Engineering*, vol. 172, p. 108158, 2023.

[35] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artificial Intelligence Review*, vol. 53, no. 8, pp. 5929–5955, 2020.

[36] R. Agharafeie, J. R. C. Ramos, J. M. Mendes, and R. Oliveira, "From shallow to deep bioprocess hybrid modeling: Advances and future perspectives," *Fermentation*, vol. 9, no. 10, p. 922, 2023.

[37] X. Cheng, Z. Guo, Y. Shen, K. Yu, and X. Gao, "Knowledge and data-driven hybrid system for modeling fuzzy wastewater treatment process," *Neural Computing and Applications*, vol. 35, no. 10, pp. 7185–7206, 2023.

[38] K. Van't Riet, "Review of measuring methods and results in nonviscous gas-liquid mass transfer in stirred vessels," *Ind Eng Chem Process Des Dev*, vol. 18, no. 3, pp. 357–364, 1979.

[39] I. Meln, W. Van Molle, M. Perea Vélez, G. Abrahamsen, K. Brusselmans, E. Calvosa, A. G. Cardillo, D. Clénet, C. Forestieri, K. V. Gernaey, M. Hoefnagel, J. B. Jørgensen, P. Lebrun, L. Natalis, B. Nilsson, V. Öppling, J. C. Pollinger, A. C. M. E. Rayat, D. Reem, M. Rubbrecht, J. Schmölder, T. Schofield, D. Smith, S. Timmins, E. von Lieres, M. Welin, and D. G. Bracewell, "Advancing regulatory dialogue: In silico models for improved vaccine biomanufacturing - an expert meeting report," *Vaccine*, vol. 56, p. 127170, 2025.