Hybrid Model Assisted Reinforcement Learning for Cell Therapy Manufacturing Process Control

Hua Zheng[1] Wei Xie[1] Keqi Wang[1] Zheng Li[2]

[1]Department of Mechanical and Industrial Engineering
Northeastern University,
[2]Genentech, Inc., South San Francisco, CA, USA

Challenges in Cell Therapy manufacturing

- **High complexity:**

Challenges in Cell Therapy manufacturing

- **High complexity:**
  - The productivity and functional identity of cell products are sensitive to cell culture conditions.

Challenges in Cell Therapy manufacturing

- **High complexity:**
  - The productivity and functional identity of cell products are sensitive to cell culture conditions.
  - Improper cultivation can not only hinder yield, but can result in heterogeneously differentiated cell populations.

Challenges in Cell Therapy manufacturing

- **High complexity:**
  - The productivity and functional identity of cell products are sensitive to cell culture conditions.
  - Improper cultivation can not only hinder yield, but can result in heterogeneously differentiated cell populations.
- **Limited data:**

Challenges in Cell Therapy manufacturing

- **High complexity:**
  - The productivity and functional identity of cell products are sensitive to cell culture conditions.
  - Improper cultivation can not only hinder yield, but can result in heterogeneously differentiated cell populations.
- **Limited data:**
  1. Lengthy analytical testing time for complex cell therapeutics

Challenges in Cell Therapy manufacturing

- **High complexity:**
  - The productivity and functional identity of cell products are sensitive to cell culture conditions.
  - Improper cultivation can not only hinder yield, but can result in heterogeneously differentiated cell populations.
- **Limited data:**
  1. Lengthy analytical testing time for complex cell therapeutics
  2. More and more personalized cell therapeutics

Challenges in Cell Therapy manufacturing

- **High complexity:**
  - The productivity and functional identity of cell products are sensitive to cell culture conditions.
  - Improper cultivation can not only hinder yield, but can result in heterogeneously differentiated cell populations.
- **Limited data:**
  1. Lengthy analytical testing time for complex cell therapeutics
  2. More and more personalized cell therapeutics
- **High variability:** seed cells can be extracted and isolated from individual patients and donors, which leads to high variability

Limitation of State-of-the-art Methods

Existing mechanistic models often ignore various sources of process stochasticity:

- batch-to-batch variation (Mockus et al., 2015)
- intracellular production fluctuations (Vasdekis et al., 2015)
- Raw material variability (Dickens et al., 2018).

Limitation of State-of-the-art Methods

Existing mechanistic models often ignore various sources of process stochasticity:

- batch-to-batch variation (Mockus et al., 2015)
- intracellular production fluctuations (Vasdekis et al., 2015)
- Raw material variability (Dickens et al., 2018).

For classical control and reinforcement learning control

- classical control strategies are often derived from deterministic mechanistic models and overlook bioprocess stochastic uncertainty & model uncertainty
- RL approaches often do not have good way to incorporate enough prior knowledge on bioprocessing mechanisms;

Hybrid-RL with Probabilistic Knowledge Graph

Driven by the critical challenges, we propose a data-driven stochastic optimization framework named "hybrid-RL".

- **KG network hybrid model** is probabilistic and mechanism-based and created to characterize the spatial-temporal causal interdependencies between critical process parameters (CPPs) and critical quality attributes (CQAs).
- **Bayesian inference** is used to derive a posterior distribution of the hybrid model.
- **Hybrid model-based Bayesian RL** (called "**hybrid-RL**") is developed to efficiently guide optimal, robust, and interpretable dynamic decision making.

The proposed Hybrid-RL framework demonstrates promising performance for cell therapy manufacturing process optimization.

Problem Statement and Bayesian RL

We model the cell therapy manufacturing process as a finite-horizon Markov decision process (MDP) specified by $(\mathcal{S}, \mathcal{A}, H, r, p)$.

The system start at an initial state $\boldsymbol{s}_1$ drawn from $p_1(\boldsymbol{s}_1)$. At any time t,

- the agent observes the state $\boldsymbol{s}_t \in \mathcal{S}$ and takes an action $\boldsymbol{a}_t \in \mathcal{A}$ from a policy $\pi_t(\boldsymbol{s}_t|\boldsymbol{a}_t)$.
- receives a reward $r_t(\boldsymbol{s}_t, \boldsymbol{a}_t) \in \mathbb{R}$.

Thus, the probabilistic model of the process trajectory $\boldsymbol{\tau} = (\boldsymbol{s}_1, \boldsymbol{a}_1, \ldots, \boldsymbol{s}_H, \boldsymbol{a}_H, \boldsymbol{s}_{H+1})$, i.e.,

$$p(\boldsymbol{\tau}|\boldsymbol{\theta}) = p(\boldsymbol{s}_1) \sum_{t=1}^{H} p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \boldsymbol{a}_t; \boldsymbol{\theta}_t),$$

Given $\boldsymbol{\theta}$, the **performance** of the policy π is evaluated via the expected accumulated reward,

$$J(\pi; \boldsymbol{\theta}) \equiv \mathbb{E}_{\boldsymbol{\tau}} \left[\sum_{t=1}^{H+1} r_t(\boldsymbol{s}_t, \boldsymbol{a}_t) \middle| \boldsymbol{\pi}, \boldsymbol{\theta} \right], \tag{1}$$

Problem Statement (Cont'd)

Latent state: Let $\boldsymbol{z}_t$ denote the latent state variables. Thus, at any time step t,

- We have observable and unobservable state $\boldsymbol{s}_t = (\boldsymbol{x}_t, \boldsymbol{z}_t)$.
- We have the likelihood of the partially observed trajectory $\boldsymbol{\tau}_x \equiv (\boldsymbol{x}_1, \boldsymbol{a}_1, \ldots, \boldsymbol{x}_H, \boldsymbol{a}_H, \boldsymbol{x}_{H+1})$ is $p(\boldsymbol{\tau}_x|\boldsymbol{\theta}) = \int \cdots \int p(\boldsymbol{\tau}|\boldsymbol{\theta}) \mathrm{d}\boldsymbol{z}_1 \cdots \mathrm{d}\boldsymbol{z}_{H+1}$.

Model uncertainty is quantified by a posterior distribution obtained by applying Bayesian rule,

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\boldsymbol{\theta}) P(\mathcal{D}|\boldsymbol{\theta}) = p(\boldsymbol{\theta}) \prod_{i=1}^{m} p\left(\boldsymbol{\tau}_x^{(i)}|\boldsymbol{\theta}\right) \tag{2}$$

where the prior $p(\boldsymbol{\theta})$ can incorporate the mechanism knowledge on the model parameters.

Objective: the optimization problem of KG hybrid model-based Bayesian RL is formulated by

$$\pi^\star = \arg\max_{\pi \in \mathcal{P}} \mathcal{J}(\pi) \tag{3}$$

with $\mathcal{P}$ representing the feasible set of decision policies and the optimization objective

$$\mathcal{J}(\pi) \equiv \mathbb{E}_{\boldsymbol{\theta} \sim p(\boldsymbol{\theta}|\mathcal{D})} [J(\pi; \boldsymbol{\theta})]$$

with (1) inner expectation in $J(\pi; \boldsymbol{\theta}) = \mathbb{E}_{\boldsymbol{\tau}}[\sum_{t=1}^{H+1} r_t(\boldsymbol{s}_t, \boldsymbol{a}_t)|\boldsymbol{\pi}, \boldsymbol{\theta}]$ accounting for inherent stochasticity; and (2) outer expectation accounting for model uncertainty.

Bioprocess Hybrid Modeling

- Given the existing ODE-based mechanistic model $\mathrm{d}\boldsymbol{s}/\mathrm{d}t = \boldsymbol{f}(\boldsymbol{s}, \boldsymbol{a}; \beta)$, we construct the hybrid model for state transition,

$$\boldsymbol{x}_{t+1} = \boldsymbol{x}_t + \Delta t \cdot \boldsymbol{f}_x(\boldsymbol{x}_t, \boldsymbol{z}_t, \boldsymbol{a}_t; \beta_t) + \boldsymbol{e}^x_{t+1}, \tag{4}$$

$$\boldsymbol{z}_{t+1} = \boldsymbol{z}_t + \Delta t \cdot \boldsymbol{f}_z(\boldsymbol{x}_t, \boldsymbol{z}_t, \boldsymbol{a}_t; \beta_t) + \boldsymbol{e}^z_{t+1}, \tag{5}$$

  where the residual terms $\boldsymbol{e}^x_{t+1} \sim \mathcal{N}(0, V^x_{t+1})$ and $\boldsymbol{e}^z_{t+1} \sim \mathcal{N}(0, V^z_{t+1})$.

- Let $g(\boldsymbol{s}_t, \boldsymbol{a}_t; \beta_t) \equiv \boldsymbol{s}_t + \Delta t \cdot \boldsymbol{f}(\boldsymbol{s}_t, \boldsymbol{a}_t; \beta_t)$. At any time step $t \in \mathcal{H}$, we have

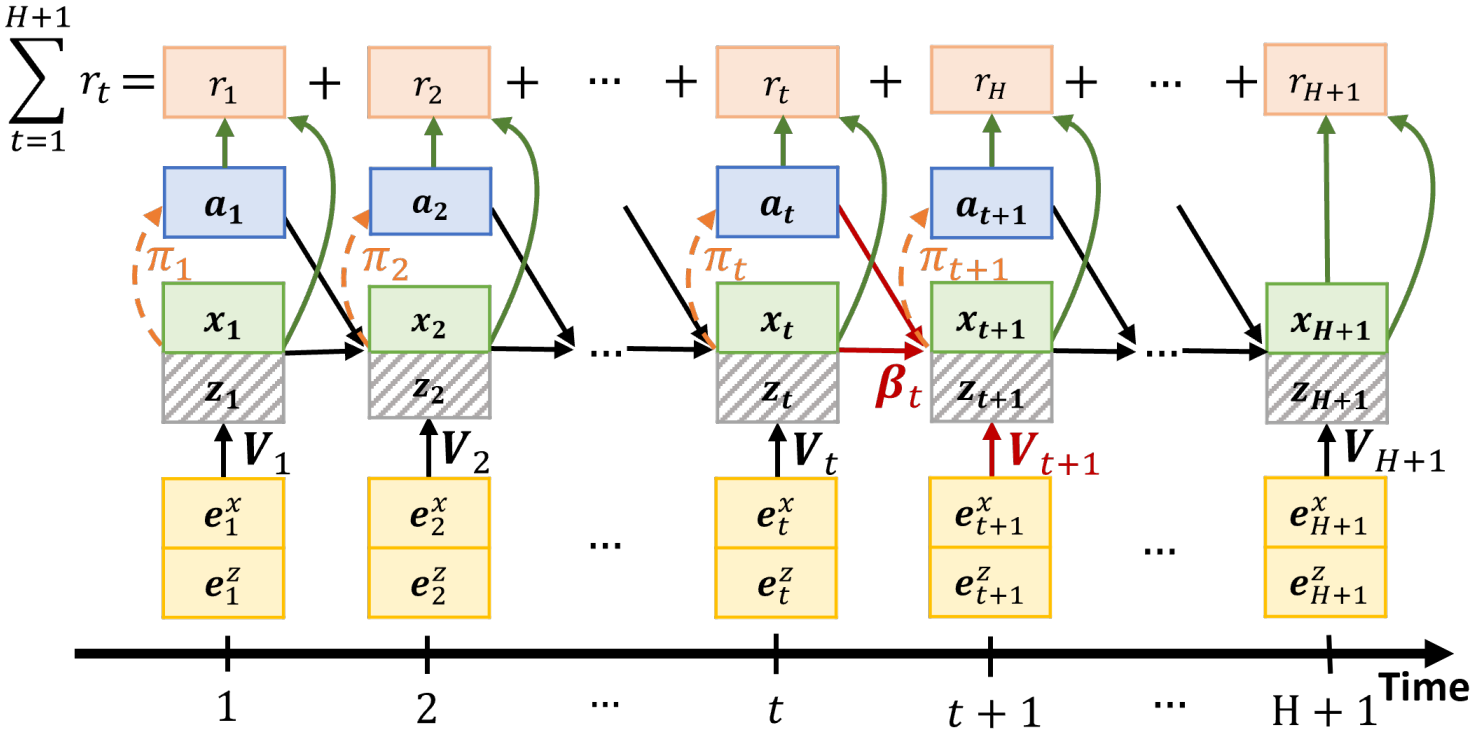
$$\boldsymbol{s}_{t+1} | \boldsymbol{s}_t, \boldsymbol{a}_t = g(\boldsymbol{s}_t, \boldsymbol{a}_t; \beta_t) + \boldsymbol{e}_{t+1} \sim \mathcal{N}\Big(g(\boldsymbol{s}_t, \boldsymbol{a}_t; \beta_t), V_{t+1}\Big) \tag{6}$$

  where $\boldsymbol{s}_t = (\boldsymbol{x}_t, \boldsymbol{z}_t)$, $\boldsymbol{e}_{t+1} = (\boldsymbol{e}^x_{t+1}, \boldsymbol{e}^z_{t+1})$, and V_{t+1} is diagonal covariance matrix with diagonal entries from V^x_{t+1} and V^z_{t+1}.

- Random mechanistic coefficients β_t account for batch-to-batch variation. $\boldsymbol{\theta}_t = \left(\boldsymbol{\mu}^\beta_t, \mathrm{vec}(\Sigma^\beta_t), \mathrm{vec}(V^x_{t+1}), \mathrm{vec}(V^z_{t+1})\right)^\top$.

Bioprocess Hybrid Modeling (Cont'd)

A Second View of Hybrid-RL: Policy-Augmented Knowledge Graph (KG)



Bayesian Inference

Inference: ABC-SMC sampling procedure for generating posterior samples from $p(\boldsymbol{\theta}|\mathcal{D})$ (derived from Toni et al. (2009); Lenormand et al. (2013); Del Moral et al. (2006)).
Main Idea: Given the observed trajectory $\boldsymbol{\tau}_x = (\boldsymbol{x}_1, \boldsymbol{a}_1, \ldots, \boldsymbol{x}_H, \boldsymbol{a}_H, \boldsymbol{x}_{H+1})$.

$$p(\boldsymbol{\theta}|\boldsymbol{\tau}_x) \propto p(\boldsymbol{\tau}_x|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{7}$$

The algorithm samples $\boldsymbol{\theta}$ and $\boldsymbol{\tau}_x^\star$ from the joint posterior:

$$p_\delta(\boldsymbol{\theta}, \boldsymbol{\tau}_x^\star|\boldsymbol{\tau}) = \frac{p(\boldsymbol{\theta})p(\boldsymbol{\tau}_x|\boldsymbol{\theta})\mathbb{1}_\delta[\boldsymbol{\tau}_x^\star]}{\int\int p(\boldsymbol{\theta})p(\boldsymbol{\tau}_x^\star|\boldsymbol{\theta})\mathbb{1}_\delta[\boldsymbol{\tau}_x^\star]d\boldsymbol{\tau}_x^\star d\boldsymbol{\theta}} \tag{8}$$

where $\mathbb{1}_\delta[\boldsymbol{\tau}_x^\star] = \mathbb{1}_\delta[d(\boldsymbol{\tau}_x, \boldsymbol{\tau}_x^\star) \leq \delta]$ is one if $d(\boldsymbol{\tau}_x, \boldsymbol{\tau}_x^\star) \leq \delta$ an zero else.

When δ is small, $p_\delta(\boldsymbol{\theta}|\boldsymbol{\tau}_x) = \int p_\delta(\boldsymbol{\theta}, \boldsymbol{\tau}_x^\star|\boldsymbol{\tau})d\boldsymbol{\tau}_x^\star$ is a good approximation to $p(\boldsymbol{\theta}|\boldsymbol{\tau}_x)$

KG Hybrid Model-based Bayesian RL

For each $t \in \mathcal{H}$, we define the state value function $V_t^\pi(\boldsymbol{s}) : \mathcal{S} \to \mathbb{R}$ and action value function $Q_t^\pi : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ as

$$V_t^\pi(\boldsymbol{s}) = \mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})} \mathbb{E}_{p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \pi_t(\boldsymbol{s}_t); \boldsymbol{\theta}_t)} \left[\sum_{\ell=t}^{H} r_\ell(\boldsymbol{s}_\ell, \pi_\ell(\boldsymbol{s}_\ell)) \middle| \boldsymbol{s}_t = \boldsymbol{s} \right]$$

$$\begin{aligned} Q_t^\pi(\boldsymbol{s}, \boldsymbol{a}) &= \mathbb{E}\left[\sum_{\ell=t}^{H} r_\ell(\boldsymbol{s}_\ell, \pi_\ell(\boldsymbol{s}_\ell)) \middle| \boldsymbol{s}_t = \boldsymbol{s}, \boldsymbol{a}_t = \boldsymbol{a} \right] \\ &= r_t(\boldsymbol{s}, \boldsymbol{a}) + \mathbb{E}\left[V_{t+1}^\pi(\boldsymbol{s}_{t+1}) \middle| \boldsymbol{s}_t = \boldsymbol{s}, \boldsymbol{a}_t = \boldsymbol{a} \right] \end{aligned} \tag{9}$$

The Bellman optimality equation (Sutton and Barto, 2018, Chapter 3.6)

$$\begin{aligned} V_t^\star(\boldsymbol{s}) &= \max_{\boldsymbol{a} \in \mathcal{A}} r_t(\boldsymbol{s}, \boldsymbol{a}) + \mathbb{E}\left[V_{t+1}^\star(\boldsymbol{s}_{t+1}) \middle| \boldsymbol{s}_t = \boldsymbol{s}, \boldsymbol{a}_t = \boldsymbol{a} \right] \\ &= \max_{\boldsymbol{a} \in \mathcal{A}} Q_t^\star(\boldsymbol{s}, \boldsymbol{a}). \end{aligned} \tag{10}$$

The optimal greedy policy (Puterman, 2014) with

$$\pi_t^\star(\boldsymbol{s}) \equiv \operatorname*{argmax}_{\boldsymbol{a} \in \mathcal{A}} Q_t^\star(\boldsymbol{s}, \boldsymbol{a}), \text{ for any } \boldsymbol{s} \in \mathcal{S}. \tag{11}$$

Bayesian Sparse Sampling (Kearns et al., 2002; Wang et al., 2005)

Input: state $\boldsymbol{s}_t$; scenario numbers B and J for estimating $\mathbb{E}_{p(\boldsymbol{\theta}|\mathcal{D})}\,\mathbb{E}_{p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\pi(\boldsymbol{s}_t);\boldsymbol{\theta}_t)}[\cdot]$; $\widehat{p}(\boldsymbol{\theta}|\mathcal{D})$ from SMC-ABC.

Output: Estimated optimal Q-function $\widehat{Q}(\boldsymbol{s},\boldsymbol{a})$

Function $\textsc{Qfun}(t,\boldsymbol{s}_t,\boldsymbol{a}_t)$**:**

- **for** $b=1,2\ldots,B$ **do**
  - **(A1)** Generate a posterior sample of model parameters, $\boldsymbol{\theta}_b \sim \widehat{p}(\boldsymbol{\theta}_t|\mathcal{D})$.
  - **for** $j=1,\ldots,J$ **do**
    - **(A2)** Sample from state transition distribution, $\boldsymbol{s}_{t+1}^{(b,j)} \sim p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\theta}_{t,b})$
    - **(A3)** $V_{t+1}\left(\boldsymbol{s}_{t+1}^{(b,j)}\right) = \textsc{Vfun}\left(t+1,\boldsymbol{s}_{t+1}^{(b,j)}\right)$
- **(A4)** $\widehat{Q}_t(\boldsymbol{s}_t,\boldsymbol{a}_t) = r_t(\boldsymbol{s}_t,\boldsymbol{a}_t) + \frac{1}{BJ}\sum_{b=1}^{B}\sum_{j=1}^{J} V_{t+1}\left(\boldsymbol{s}_{t+1}^{(b,j)}\right)$.
- **return** $\hat{Q}_t(\boldsymbol{s}_t,\boldsymbol{a}_t)$.

Function $\textsc{Vfun}(t,\boldsymbol{s}_t)$**:**

- **if** $t=H+1$ **then**
  - **return** $r_{H+1}(\boldsymbol{s}_{H+1})$;
- **for** $\boldsymbol{a}_t \in \mathcal{A}$ **do**
  - **for** $b=1,2\ldots,B$ **do**
    - **(B1)** Generate a posterior sample of model parameters $\boldsymbol{\theta}_b \sim \widehat{p}(\boldsymbol{\theta}_t|\mathcal{D})$
    - **for** $j=1,\ldots,J$ **do**
      - **(B2)** Sample from state transition $\boldsymbol{s}_{t+1}^{(b,j)} \sim p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\theta}_{t,b})$
      - **(B3)** $V_{t+1}\left(\boldsymbol{s}_{t+1}^{(b,j)}\right) = \textsc{Vfun}\left(t+1,\boldsymbol{s}_{t+1}^{(b,j)}\right)$
  - **(B4)** Estimate $\widehat{Q}_t(\boldsymbol{s}_t,\boldsymbol{a}_t) = r_t(\boldsymbol{s}_t,\boldsymbol{a}_t) + \frac{1}{BJ}\sum_{b=1}^{B}\sum_{j=1}^{J} V_{t+1}\left(\boldsymbol{s}_{t+1}^{(b,j)}\right)$
- **(B5)** $\widehat{V}_t(\boldsymbol{s}_t) = \max_{\boldsymbol{a}_t\in\mathcal{A}} \widehat{Q}_t(\boldsymbol{s}_t,\boldsymbol{a}_t)$ as in (10)

Cell Therapy Manufacturing Case Study

Mechanistic Model: Glen et al. (2018) developed an ODE-based mechanistic model describing the dynamics of an unidentified autocrine growth inhibitor accumulation and its impact on the erythroblast cell production process. We extend this model to a two phases: growth and stationary phases with index $p = 1, 2$.

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = r_p^g \rho \left(1 - \left(1 + e^{(k_p^s(k_p^c - I))}\right)^{-1}\right), \tag{12}$$

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \frac{\mathrm{d}\rho}{\mathrm{d}t} - r_p^d I, \tag{13}$$

- ρ_t and I_t represent the cell density and the inhibitor concentration at t.
- The kinetic coefficients r_p^g, k_p^s, k_p^c and r_p^d denote the cell growth rate, inhibitor sensitivity, inhibitor threshold, and inhibitor decay. The phase transition occurs at $T_\star = 18$ hour.

Simulator: based on (12)-(13), we develop a simulator by including various source of uncertainty

$$\mathrm{d}\rho = r_p^g \rho \left(1 - \left(1 + e^{(k_p^s(k_p^c - I))}\right)^{-1}\right)\mathrm{d}t + \sigma_n \mathrm{d}W \tag{14}$$

$$\mathrm{d}I = \mathrm{d}\rho - r_p^d I \mathrm{d}t + \sigma_n \mathrm{d}W \tag{15}$$

with (1) **random initial values** $\rho_1 \sim \mathcal{N}(\mu_\rho, \sigma_\rho^2)$ and $I_1 = 0$, (2) **batch-to-batch variation** $r_p^g \sim \mathcal{N}\left(\mu_p^g, (\sigma_p^g)^2\right)$ with $p = 1, 2$, and **measurement error** $\rho_t \leftarrow \rho_t + e_m$ with $e_m \simeq \mathcal{N}(0, \sigma_m^2)$.

Prediction Error (MAE) of Cell Density

Assessment of the long-term prediction performance (mean absolute error) of the KG hybrid model and the ODE mechanistic model fitted by LS method (LS-ODE).

Both models were fitted by "real-world" historical trajectories with the size $m = 3, 6, 20$. We evaluate performance based on $r = 30$ macro-replications.

Noise Level		h (hrs)	Hybrid			LS-ODE		
b2b	noise		$m = 3$	$m = 6$	$m = 20$	$m = 3$	$m = 6$	$m = 20$
high	$\sigma_n = 0.01$	3	0.12 ± 0.05	0.09 ± 0.03	0.06 ± 0.02	0.41 ± 0.19	0.59 ± 0.30	0.44 ± 0.22
		18	0.60 ± 0.17	0.48 ± 0.10	0.26 ± 0.07	0.74 ± 0.15	0.57 ± 0.25	0.49 ± 0.23
		30	0.59 ± 0.16	0.40 ± 0.11	0.22 ± 0.06	0.65 ± 0.24	0.70 ± 0.36	0.84 ± 0.65
high	$\sigma_n = 0.03$	3	0.21 ± 0.05	0.14 ± 0.03	0.08 ± 0.03	0.37 ± 0.20	0.40 ± 0.19	0.36 ± 0.23
		18	1.07 ± 0.22	0.82 ± 0.15	0.48 ± 0.12	1.11 ± 0.28	0.93 ± 0.31	0.83 ± 0.34
		30	1.11 ± 0.24	0.74 ± 0.16	0.44 ± 0.12	1.57 ± 0.76	1.09 ± 0.41	0.93 ± 0.45
low	$\sigma_n = 0.01$	3	0.10 ± 0.03	0.07 ± 0.02	0.04 ± 0.01	0.38 ± 0.23	0.54 ± 0.26	0.38 ± 0.20
		18	0.48 ± 0.12	0.38 ± 0.09	0.27 ± 0.06	0.43 ± 0.13	0.35 ± 0.12	0.28 ± 0.11
		30	0.47 ± 0.11	0.30 ± 0.08	0.16 ± 0.04	0.45 ± 0.11	0.52 ± 0.23	0.32 ± 0.15
low	$\sigma_n = 0.03$	3	0.18 ± 0.06	0.13 ± 0.03	0.04 ± 0.01	0.68 ± 0.34	0.46 ± 0.23	0.31 ± 0.18
		18	1.00 ± 0.20	0.69 ± 0.15	0.27 ± 0.06	1.27 ± 0.36	1.47 ± 1.41	0.55 ± 0.16
		30	1.04 ± 0.28	0.65 ± 0.17	0.16 ± 0.04	1.40 ± 0.48	1.61 ± 1.40	0.66 ± 0.20

Remark: b2b is batch-to-batch variation and noise is the process noise.

Medium Full Exchange Decision Making

Background: Medium exchange is an essential element of successful long-term cell culture. Culture medium is exchanged to supply new nutrients and to eliminate waste products produced by the cells.

Goal: finding the optimal time to fully exchange the medium with fresh medium.

State: is defined as cell density and inhibitor $\boldsymbol{s}_t = (\rho_t, I_t)$.

Action: $a_t = 0$ denoting the full exchange of medium at step t; $a_t = 1$, otherwise. $I_t = a_t I_t$ represent the post-exchanged concentration of inhibitor

Reward:

- Total operational cost: $C(T, M) = C_t T + C_m M$,
- Reward function is defined by cell yield per cost — the efficiency of the system during the T hours (H time steps) cell culture (Glen et al., 2018):

$$r_t = 0 \quad \text{with} \quad 0 \leq t \leq H$$

$$r_{H+1}(\boldsymbol{s}_{H+1}, a_{H+1} = \text{“Harvest”}) = \frac{M(\rho_T - \rho_0)}{C(T, M)}$$

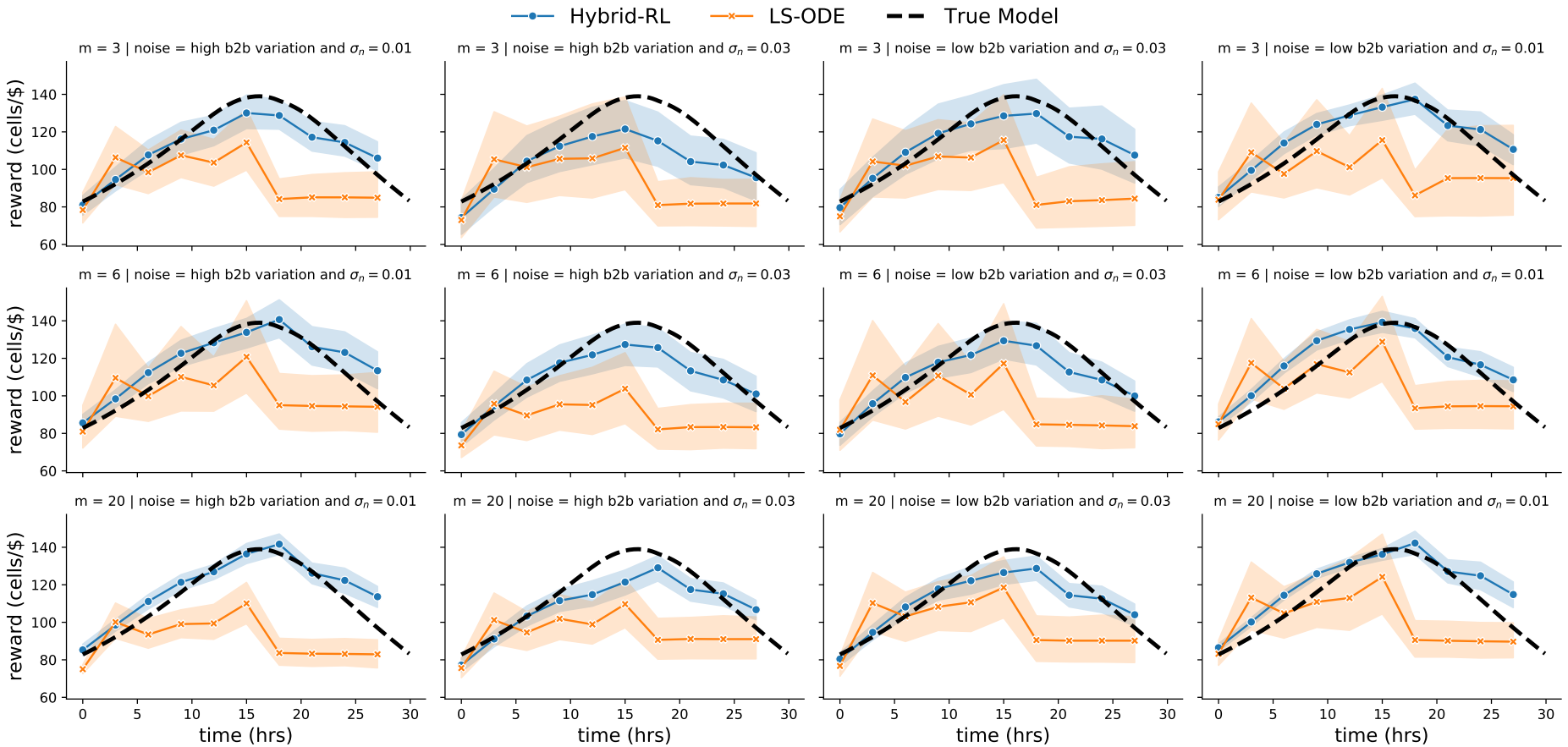
where ρ_T represents the cell density at the T-th hour.

Decision Epochs: The medium is fully exchanged up to one time in decision hours $\{0, 3, \ldots, 27\}$.

Medium Exchange Cost Efficiency

Performance of hybrid-RL and LS-ODE in 30 macro-replications.

- The validated models are used to optimize the media exchange time for cells to be produced with optimal cost efficiency at a given production scale (100L).
- The number of cells produced per dollar (y-axis) for a given time point (x-axis) of media exchange are calculated for \$150/hr operating time cost and \$10/L of consumable cost



Cell Culture Expansion Scheduling

- Each expansion, the original batch is scaled up to a n times larger cell culture vessel filling with fresh medium.
- Cell density ρ and the concentration of inhibitor I decrease to $1/n$ of original batch immediately after each scale-up.
- The reward function is then defined by the difference between revenue and cost as,

$$
\begin{aligned}
r_t &= 0 \text{ with } 0 \leq t \leq H \\
r_{H+1}(\boldsymbol{s}_{H+1}, a_{H+1} &= \text{“Harvest”}) \\
&= K(\rho_T, \xi, n) - C(T, M)
\end{aligned}
$$

where the revenue $K(\rho_T, \xi, n) = P_c \times \rho_T \times n^{\xi}$

Noise Level		Hybrid-RL			LS-ODE		
b2b	noise	$m = 3$	$m = 6$	$m = 20$	$m = 3$	$m = 6$	$m = 20$
high	$\sigma_n = 0.01$	7317.24 (352.37)	7588.15 (322.53)	7892.84 (69.65)	5677.62 (389.24)	5944.25 (403.52)	6030.83 (399.61)
high	$\sigma_n = 0.03$	6888.86 (693.07)	7266.23 (393.66)	7689.45 (143.11)	-2259.01 (704.82)	1026.60 (484.50)	2454.22 (262.15)
low	$\sigma_n = 0.01$	7800.60 (151.01)	7955.38 (75.78)	8035.71 (52.48)	6115.31 (413.23)	6193.70 (381.83)	6417.39 (389.06)
low	$\sigma_n = 0.03$	7414.34 (334.41)	7572.99 (329.62)	7974.35 (76.15)	5978.52 (389.96)	6126.60 (393.61)	6225.02 (395.50)

Thank you!

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006.

Jason Dickens, Sarwat Khattak, Thomas E Matthews, Dave Kolwyck, and Kelly Wiltberger. Biopharmaceutical raw material variation and control. *Current opinion in chemical engineering*, 22:236–243, 2018.

Katie E Glen, Elizabeth A Cheeseman, Adrian J Stacey, and Robert J Thomas. A mechanistic model of erythroblast growth inhibition providing a framework for optimisation of cell therapy manufacturing. *Biochemical Engineering Journal*, 133: 28–38, 2018.

Michael Kearns, Yishay Mansour, and Andrew Y Ng. A sparse sampling algorithm for near-optimal planning in large markov decision processes. *Machine learning*, 49(2):193–208, 2002.

Maxime Lenormand, Franck Jabot, and Guillaume Deffuant. Adaptive approximate bayesian computation for complex models. *Computational Statistics*, 28(6):2777–2796, 2013.

Linas Mockus, John J Peterson, Jose Miguel Lainez, and Gintaras V Reklaitis. Batch-to-batch variation: a key component for modeling chemical manufacturing processes. *Organic Process Research & Development*, 19(8):908–914, 2015.

Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.

Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018. ISBN 0262039249.

Tina Toni, David Welch, Natalja Strelkowa, Andreas Ipsen, and Michael PH Stumpf. Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems. *Journal of the Royal Society Interface*, 6(31): 187–202, 2009.

AE Vasdekis, Andrew M Silverman, and Gregory Stephanopoulos. Origins of cell-to-cell bioprocessing diversity and implications of the extracellular environment revealed at the single-cell level. *Scientific Reports*, 5(1):1–7, 2015.

Tao Wang, Daniel Lizotte, Michael Bowling, and Dale Schuurmans. Bayesian sparse sampling for on-line reward optimization. In *Proceedings of the 22nd international conference on Machine learning*, pages 956–963, 2005.