Probabilistic Modeling for Optimization of Bioreactors using Reinforcement Learning with Active Inference

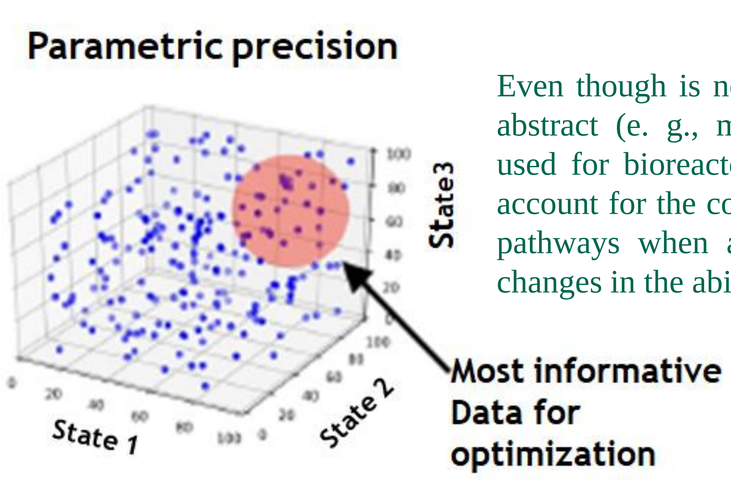
Ernesto C. Martínez, Jong Woo Kim, Tilman Barz, M. Nicolás Cruz B.

Technische Universität Berlin, Chair of Bioprocess Engineering, KIWI-biolab
www.bioprocess.tu-berlin.de
www.kiwi-biolab.de

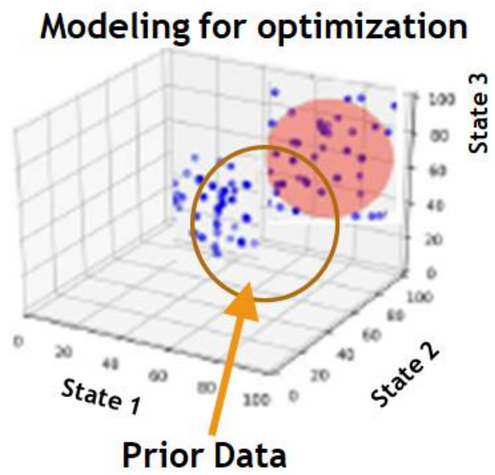
Institute of Biotechnology
Berlin Institute of Technology
Bioprocess Engineering

Motivation and Scope

With **low-complexity imperfect** models, bioreactor optimization is often **sub-optimal**. Why?



Even though is not widely accepted nor recognized, abstract (e. g., macroscopic or cybernetic) models used for bioreactor optimization are too shallow to account for the complexity of switching in metabolic pathways when a microorganism is responding to changes in the abiotic conditions



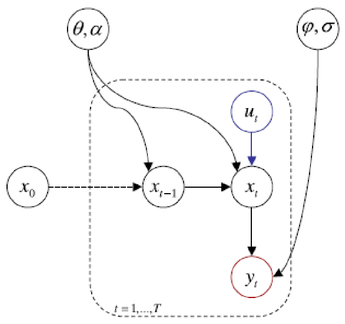
A challenge in *modeling for optimization* is how effective bioreactor models can be learned from designed experiments given (i) the rich complexity of profiling operating conditions, and (ii) the circular dependence of model learning and information content of sampled data, which often leads to suboptimal performance and low reproducibility.

Methodology and Results

Probabilistic (causal) models of bioreactors

A probabilistic (causal) model of a bioreactor is defined by a joint probability distribution over the following set of stochastic variables:

- $x; y$: the $n\ x\ n_t$ hidden states time-series; the $p\ x\ n_t$ observations (sampled data),
- u: the $n_u\ x\ n_t$ manipulated (controlled) inputs time-series,
- $\theta; \varphi$: the $n_\theta x\ 1$ evolution parameters; the $n_\varphi x\ 1$ observation parameters,
- α: the state noise precision (structural errors),
- σ: the measurement noise precision (analytical and sensor calibration errors).



These variables are assumed to follow the (hidden) **state evolution and observation equations**:

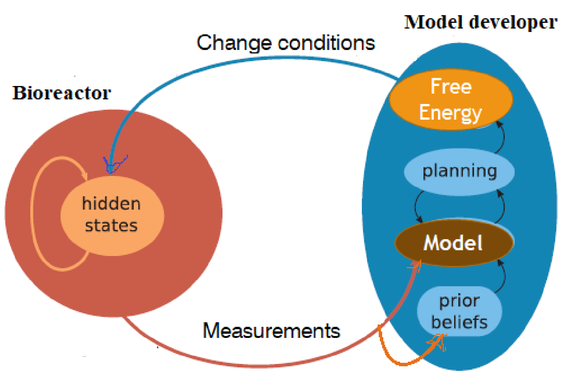
$$x_t = f(x_{t-1}, \theta, u_{t-1}) + \eta_t;\ \eta_t = N(0, \alpha^{-1}\boldsymbol{I})$$
$$y_t = g(x_t, \varphi) + \varepsilon_t;\quad \varepsilon_t = N(0, \sigma^{-1}\boldsymbol{I})$$

The **probabilistic model** m of a bioreactor is completely specified by the (initial) *Gaussian* prior distributions for its parameters θ, φ, and the *Gamma* priors for the precision hyper-parameters α, σ.

Active inference (Goal-directed sampling)

Active inference proposes that the modeler´s goal or intent are encoded in the probabilistic model as a prior preference for desired observations (e. g., higher biomass productivity or protein expression).

Active Inference = Active Learning + Variational Inference



Probabilistic model learning is posed as the maximization of a free-energy lower bound $F(q)$ functional for the model evidence with respect to an approximate density q_{Φ}

$$F(q) = \langle \ln p(\Phi|m) + p(\Phi|y,m) - p(\Phi) \rangle_q$$
$$ELBO = \ln p(y|m) - \mathcal{D}_{KL}\big(p(\Phi); p(\Phi|y,m)\big)$$

where $\mathcal{D}_{KL}$ is the **Kullback-Leibler divergence**.

Adaptive optimization-oriented redesign

Inputs: T, K, x_0, prior $q(\boldsymbol{\Phi})$, state evolution and observation functions f, g

▷ **For** $t = 1$ to $T - 1$

Infer current state $\hat{x}_t$ using $u^*_{t-1}, \hat{x}_{t-1}$ and Thompson *Sampling* of prior $q(\boldsymbol{\Phi})$

▷ **For** $k = 1$ to K

$\hat{q}(\boldsymbol{\Phi_k}) = q(\boldsymbol{\Phi})$

While $t < T$ (Forward Pass)

Thompson *Sampling* of the prior $\hat{q}(\boldsymbol{\Phi_k})$: $\phi_k = \boldsymbol{TS}[\hat{q}(\boldsymbol{\Phi_k})]$

$u_t^k = argmax_{u\in\Omega}\ r_{t+1}(\hat{y}_{t+1}|\phi_k, \hat{x}_t)$

Simulate redesign using u_t^k and predict $r_{t+1}\left(\widetilde{y_{t+1}^k}\middle| u_t^k, \hat{x}_t\right)$

Update prior: $\hat{q}(\boldsymbol{\Phi_k}) \leftarrow \hat{q}(\boldsymbol{\Phi_k}|u_t^k, y_{t+1}^k)$ using (u_t^k, y_{t+1}^k)

Accumulate reward: $\mathcal{R}_k = \mathcal{R}_k + r_{t+1}\left(\widetilde{y_{t+1}^k}\middle| u_t^k\right)$

End while

Define the policy: $\pi_t^k = \{u_t^k, \dots, u_{T-1}^k\}$ with its corresponding $\mathcal{R}_k$

▷ **End for**

Rank policies $\pi_t^k, k = 1, .., K$, using $\mathcal{R}_k$

Select the best policy $\pi_t^* = \{u_t^*, \dots, u_{T-1}^*\}$ with the highest $\mathcal{R}_k$

Redesign **the experiment** using only u_t^* and measure y_{t+1} at $t + 1$

Update prior: $q(\boldsymbol{\Phi}) \leftarrow q(\boldsymbol{\Phi}|u_t^*, y_{t+1})$ using experimental data (u_t^*, y_{t+1})

▷ **End for**

Outputs: $\pi^* = \{u_1^*, \dots, u_{T-1}^*\}$, $\boldsymbol{y} = \{y_1, \dots, y_T\}$, $\boldsymbol{r} = (r_1, \dots, r_T)$, $q(\boldsymbol{\Phi})$

Reinforcement learning for online redesign

Let $z_{t:T}$ denote a sequence of variables through time, $z_{t:T} = \{z_t, .., z_T\}$, and **let define a policy** Π **as a way of behaving** over time**:**

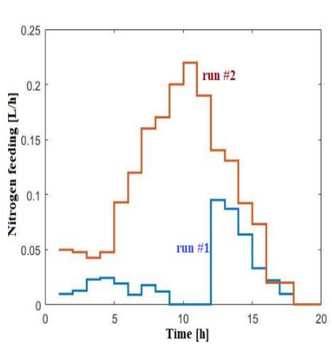
$u \leftarrow \boldsymbol{\Pi}(x_t)$, (the map **Π** here is a probabilistic assignment from states to actions).

Applying recursively the policy Π defines **sequence of actions** $\pi = \{u_t, \dots, u_{T-1}\}$. In "modeling for optimization," the specific aim is to *minimize the free energy of the expected future* $\tilde{F}_\pi$, which is defined as:

$$\tilde{F}_\pi = \mathcal{D}_{KL}\big(q(y_{t:T}, x_{t:T}, \boldsymbol{\Phi}|\pi) \| p^*(y_{t:T}, x_{t:T}, \boldsymbol{\Phi})\big);\ \boldsymbol{\Phi} = (\theta, \varphi, \alpha, \sigma)$$

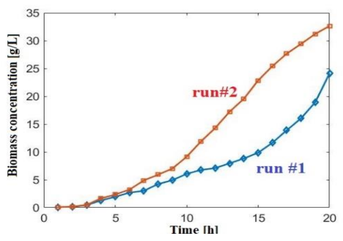
where $q(y_{t:T}, x_{t:T}, \Phi|\pi)$ models the probability distribution for future trajectories in a dynamic experiment under policy π and $p^*(y_{t:T}, x_{t:T}, \boldsymbol{\Phi})$ defines the joint probability distribution for the optimal trajectory of the hidden states, model parameters and preferred observations. Thus, when $\tilde{F}_\pi$ is driven to zero, the policy π becomes the (probabilistic) optimal policy. Notice that by minimizing $\tilde{F}_\pi$, the surprise $-\ln\ p(y_{t:T}|m)$ is also minimized, which in turns maximize the Bayesian model evidence.

Results for the Baker's yeast production process



Units	Run #1 μ	Run #1 σ	Run #2 μ	Run #2 σ	Run #3 μ	Run #3 σ
gX/gG	0.5431	0.250	0.6232	0.112	0.7012	0.010
gX/gG	0.0612	0.020	0.0578	0.011	0.8510	0.006
gX/gE	0.8929	0.085	0.8500	0.005	0.8091	0.002
gE/gG	0.2647	0.080	0.2450	0.009	0.2350	0.002
gA/gX/gG	0.2589	0.080	0.0189	0.005	0.2162	0.037
gX/gN	1.0150	0.150	0.9733	0.088	0.8817	0.044
gG/gX/h	0.4445	0.125	0.4210	0.039	0.4198	0.028
gG/gX/h	2.5364	0.200	2.6472	0.145	2.7116	0.009
gN/gX/h	1.1903	0.150	1.2279	0.067	1.2009	0.011
gG/L	0.1524	0.030	0.1208	0.025	0.0989	0.010
gE/L	3.1817	0.050	3.2011	0.443	2.9442	0.031
gN/L	2.9370	0.050	2.9674	0.632	3.3598	0.278
gA/gX/L	9.0014	2.000	9.4569	1.227	10.106	0.583
gA/gX/L	5.5981	0.500	5.8919	0.389	6.2991	0.267
gA/gX/L	5.5737	0.210	6.131	0.201	5.8404	0.113

Modeling Run #	Biomass [g/L]
1	24.11
2	32.64
3	31.58
Richelle et al. 2014	32.00



Significance and Concluding Remarks

A novel probabilistic method for modeling the dynamic behavior of bioreactors in the most profitable region of operating conditions is proposed. Based on simulation data, a dynamic experiment is redesigned online through active inference. Reinforcement learning is used to maximize the Bayesian model evidence, that is, to minimize surprise.

- ✓ Probabilistic (causal) models of bioreactors are learned by biasing data gathering using the **Free Energy of the Expected Future**.
- ✓ Reinforcement learning following an MPC-approach is used to **combine planning and control for online experiment redesign**.
- ✓ Bayesian Variational Analysis methods are applied for state inference and probabilistic parameter estimation.
- ✓ Simulation data is used to **learn a redesign policy** for adaptive experimental design.

Acknowledgements / References

This work was supported by the German Federal Ministry of Education and Research through the Program "International Future Labs for Artificial Intelligence" (Grant number 01DD20002A).