Learning probabilistic models of biological systems using active inference with belief propagation

Ernesto C. Martínez

INGAR (CONICET-UTN), Avellaneda 3657, 3000 Santa Fe, Argentina

Abstract. In this work, the normative framework of active inference is integrated with belief propagation for inverting a probabilistic causal model using data generated from planned interactions between a Bayesian modeling agent and a biological system. Thompson sampling of parameter distributions is used to estimate the free energy of the expected future when beliefs about beliefs are rolled over a planning horizon. Learning a probabilistic model for maximizing biomass production in the well-known Baker's yeast example is used as an example. The prior parameter distributions in the system model of a fed-batch cultivation are updated as new observations are obtained. Planned action sequences aim to excite the yeast metabolism by introducing changes in the feed rate of two nutrients (glucose and nitrogen). Results obtained demonstrate that by maximizing the model evidence, the proposed approach constraints biological system dynamics to relevant trajectories for improved parametric precision in the preferred region of physiological states that favor biomass productivity.

Keywords: Active inference, Bayesian inference, Biological systems, Probabilistic modeling, Reinforcement learning.

1 Introduction

Even though is not yet accepted nor recognized, abstract (e. g., macroscopic or cybernetic) models used to describe the response of biological systems are too shallow to account for the full complexity of switching in metabolic pathways when responding to changes in their abiotic conditions [4]. A challenge in gathering informative data is how a causal probabilistic model can be learned from designed experiments given (i) the rich complexity of time-varying environmental conditions, and (ii) the circular dependence of model learning and information content of sampled data, which may lead to inaccurate predictions of a micro-organism response to environmental stimuli [10]. As most models of biological systems are not a veridical representation due to regulatory mechanisms in its metabolism, it is infeasible to achieve parametric precision comprehensively, let alone design optimally informative experiments for this objective [7, 8]. Probabilistic causal models are better prepared to deal with the system-model structure mismatch and unobserved hidden states [2, 3]. In this work, we illustrate how ideas from active inference [9, 15] can be integrated with belief propagation to unify goal-oriented and information-seeking objectives in modeling biological systems using the variational free energy of the expected future over a sequence of

interactions between the modeler that modifies the abiotic conditions and the biological system that respond to purposefully designed stimuli. Active inference replaces the value functions in reinforcement learning with functionals of (Bayesian) beliefs and modeler´s preferences, in the form of an expected (variational) free energy. To plan using beliefs about beliefs some short of sophistication is needed [5].

1.1 The Bayesian modeling agent-biological system interaction cycle

In this work, learning of a probabilistic causal model is based on data gathered in a sequence of planed interactions between a Bayesian modeling agent and a biological system. The modeling agent has *a priori* belief about the causal response of the biological system to external stimuli as shown in Fig. 1. The received reward measures the goodness of the system response depending on the agent preferences. Based on its beliefs about the parameter distributions and model structure, including the hidden states that causally explain the expected observations, the agent only takes the first action of the planned sequence to maximize the information content of probable future states in the most rewarded state trajectories. As soon as the system responds to the actions taken, beliefs about its behavior are first updated by the agent. Then, a new sequence of actions is designed to maximize information gain by rolling beliefs about beliefs into the future.

The Bayesian modeling agent plans future interactions based on its beliefs about the counterfactual consequences of alternative actions for hidden physiological states and the updated beliefs (parameter distributions) resulting from distributions for hidden states in the simulated sequence of state transitions and predicted rewards. This recursive form of belief propagation using a probabilistic causal model of the biological system implements effectively a deep tree search over course of actions and expected system response in future interactions to bias data gathering for model building in the most preferred states as measured by the rewards predicted using the model. Key to propagating beliefs about belief is efficient posterior sampling.

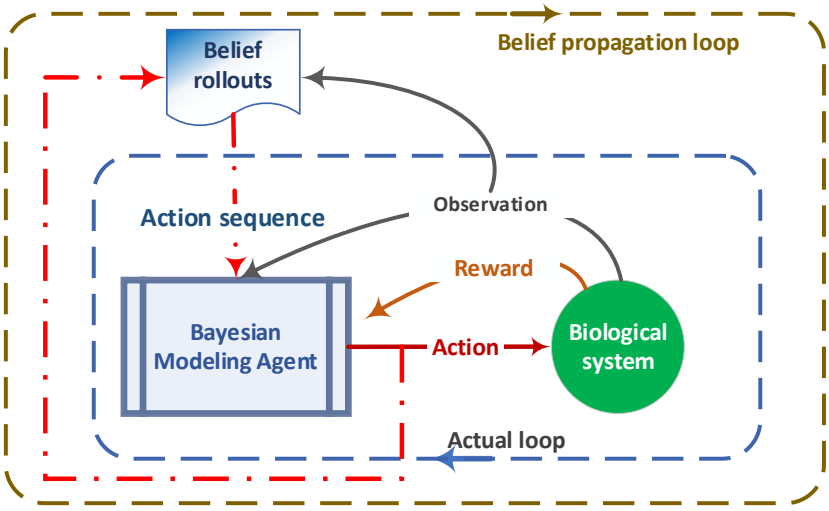


Fig. 1. Bayesian modeling cycle based on counterfactual belief propagation.

2 Probabilistic causal models

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

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

From the sequence of interactions between the modeling agent and the biological system under study, it is assumed that these variables follow the equations:

$$\begin{aligned} x_t &= f(x_{t-1}, \theta, u_{t-1}) + \eta_t;\ \eta_t = N(0, \alpha^{-1}\boldsymbol{I}), && \text{(hidden) state evolution,} \\ y_t &= g(x_t, \varphi) + \varepsilon_t;\quad \varepsilon_t = N(0, \sigma^{-1}\boldsymbol{I}), && \text{observation (response).} \end{aligned} \tag{1}$$

where *f* (resp. *g*) is the first-principles model (observation model), and η_t (observation ε_t) is the state (resp. measurement) modeling errors (noise). A probabilistic model *m* of a bioreactor is completed by specifying the (initial) Gaussian prior distributions for its parameters θ, φ. Also, Gamma distribution priors are defined for the precision hyperparameters α, σ. Given these priors, the left part of Eq. (1) induces a (so-called semi-Markovian process) prior density on the trajectory of hidden states *x*. Similarly, the right part of Eq. (1) yields a likelihood function which measures how plausible an observation *y* is when the biological system responds to an environmental stimuli u_{t-1} at time *t*. Uncertainties from noisy observations and model imperfections are thus taken explicitly into account by the probabilistic model.

In the variational Bayesian framework, model identification (or inversion) entails the estimation of the marginal likelihood or evidence of a bioreactor model, that is a probabilistic description of the main (causal) metabolic mechanisms by which sampled data are generated. Probabilistic Bayesian treatment of an experiment dataset makes full usage of prior assumptions regarding the statistical distributions for initial conditions, evolution/observation parameters and state/measurement noise [2, 3]. Inverting a probabilistic model *m* requires approximating the conditional density $p(\xi|y,m)$ of the unknown hidden states and parameters $\xi = \{x, x_0, \theta, \varphi, \alpha, \sigma\}$ given a data set of sampled measurements *y* and computing the model evidence $p(y|m)$. Nonlinearities in the probabilistic model prevent exact analytical solutions to the model inversion problem which can approximately be solved using Bayesian variational approaches such as active inference.

The plausibility of a model parameter vector $\Phi = (\theta, \varphi, \alpha, \sigma)$ given a model structure *m* having a new dataset $\mathcal{D}$ can be expressed by the **Bayes´ Rule**:

$$P(\Phi|m,\mathcal{D}) = \frac{P(\mathcal{D}|\Phi,m)\,P(\Phi|m)}{P(\mathcal{D}|m)} \tag{2}$$

As it is shown in Fig. 2, based on a new dataset $\mathcal{D}$, a prior distribution for model parameters is update into a posterior distribution for Φ. If the new dataset is highly informative the posterior density is significantly changed due to the bias introduced. However, if the incoming data is not so surprising, the posterior update is much less significant. The key question is how purposefully excite the metabolic response of a biological system to gain the most from resulting data?

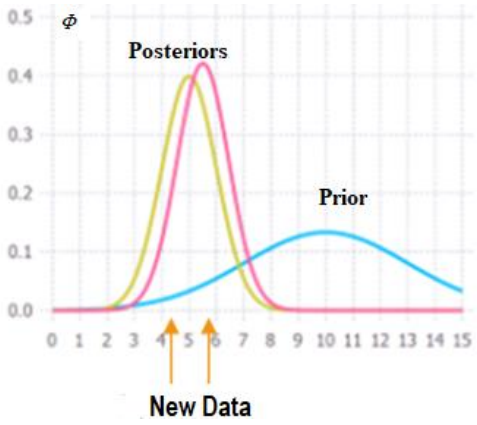


Fig. 2. Posterior update of a parameter distribution upon new data and their surprise contents.

One of the core problems of model inversion in a Bayesian setting is to approximate difficult-to-compute posterior probability densities for model parameters from their priors. In this work, for model inversion is carried out using Variational Inference [1]. Rather than resorting to sampling, the main idea behind variational inference is to use optimization. First, we posit a family of approximate densities $q \in \mathcal{Q}$. Then, we try to find the member of that family that minimizes the *Kullback-Leibler* divergence or distance $\mathcal{D}_{KL}$ to the exact posterior distribution when new data y is accounted for by solving:

$$q^*(\Phi) = \frac{\arg\, min}{q(\Phi)\in\mathcal{Q}}\ \mathcal{D}_{KL}\big(q(\Phi); p(\Phi|y,m)\big); \tag{3}$$

Thus, we approximate the posterior with an optimized member of the family $q^*(\Phi)$. Variational inference thus turns the inference problem into an optimization problem, and the reach of the family $\mathcal{Q}$ manages the complexity of this functional approximation of the posterior. One of the key ideas behind variational inference is to choose $\mathcal{Q}$ to be flexible enough to capture a density close to the exact posterior $p(\Phi|y,m)$, but simple enough to be found by efficient optimization.

When inverting realistic probabilistic models of biological systems, nonlinearities in the likelihood function generally induce posterior densities that are not in the conjugate family of their priors. The Laplace, also known as *mean-field*, approximation is a useful approach, which can finesse this problem by reducing the set of sufficient

statistics of the approximate posterior density to its first two moments, mean and variance. This means that that both prior and approximate marginal posterior density are assumed to follow a Gaussian density, except for the precision hyperparameters α and σ, which are assumed to have Gamma posterior densities. The main assumption behind the Laplace approximation is that model parameters are independent:

$$q_i(\Phi)=\prod_1^k q_i\ (\theta_i) \quad \textbf{(4)}$$

Thus, the variational Bayesian (VB) update of parameter posteriors reduce to a regularized Gauss-Newton optimization scheme [2, 3]. This dramatically decreases the computational complexity of solving the optimization problem in Eq. 2. The second-order moments of the approximate posterior densities are then simply related to the curvature of Kullback-Leibler distances between priors and candidate posteriors (distributions).

3 Active inference

Active inference is a normative theory that unifies observation, external stimulus, and model learning under a single imperative—the minimization of the variational free energy [5, 9]. More specifically, probabilistic model learning is posed as the maximization of a free energy lower bound $F(q)$ for the model evidence with respect to an approximate density $q(\xi)$:

$$F(q) = \langle \ln p(\xi|m) + p(\xi|y,m) - q(\xi) \rangle_q = \ln p(y|m) - \mathcal{D}_{KL}\big(q(\xi); p(\xi|y,m)\big) \quad \textbf{(5)}$$

where $\mathcal{D}_{KL}$ is the Kullback-Leibler divergence between two distributions and the expectation $\langle \circ \rangle_q$ is taken under the approximate posterior distribution q. As can be deduced from Eq. (2), maximizing the functional $F(q)$ with respect to q drives the Kullback-Leibler divergence between $p(\xi)$ and the exact posterior $p(\xi|y,m)$ to zero. The reader is referred to the work of Daunizeau et al. [2] for methodological details and the references therein described VBA Toolbox for variational Bayesian analysis.

Active inference proposes that the modeler´s goal or intent are encoded in the probabilistic model as prior preferences for favourable observations (e. g., higher biomass productivity or protein expression). Thus, active inference demands that parameter distributions in the probabilistic model are based on data sampled from favourable operating conditions. This can be achieved by means of online replanning of interaction over a horizon aiming at minimizing the free energy of the expected future [9], which corresponds to the trajectory of hidden states that is expected to occur from applying the optimal sequence of actions (sampling times and manipulated inputs) that maximise the model evidence in the certain region of environmental conditions.

Let $z_{t:T}$ denote a sequence of variables through time, $z_{t:T} = \{z_t, \ldots, z_T\}$, and let define a policy as a sequence of actions $\pi = \{u_t, \ldots, u_{T-1}\}$. In probabilistic modeling of biological systems the specific aim is to minimize the free energy of the expected future $\tilde{F}_\pi$, which is defined as:

$$\tilde{F}_{\pi} = \mathcal{D}_{KL}\left(q(y_{t:T}, x_{t:T}, \boldsymbol{\Phi}|\pi) \| p^{*}(y_{t:T}, x_{t:T}, \boldsymbol{\Phi})\right); \boldsymbol{\Phi} = (\theta, \varphi, \alpha, \sigma) \quad \textbf{(6)}$$

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

Active inference´s main argument is that the modeler´s goal or intent should be encoded in the probabilistic model as a prior preference for desired observations (e. g., higher biomass productivity or protein expression). Thus, active inference demands that parameter distributions in the probabilistic model are based on data sampled from favourable physiological conditions.

4 Belief propagation through posterior Thompson sampling

Inspired by posterior (Thompson) sampling for multi-armed bandits [12, 13], a simple belief propagation algorithm which uses data from previous interactions to plan a sequence of future outcomes (observations and corresponding rewards) from optimized actions that maximize model evidence is proposed (see Fig. 3). At a given discrete time *t*, let´s assume a priori distributions $q(\boldsymbol{\Phi})$ for model parameters. Using a rolling horizon *H* for belief propagation based on Thompson sampling to assess the effect of incorporating different realizations of simulation data, the estimated optimal action is taken and the priors for model parameters are updated at each time step. A reward function that accounts for the modeling agent preferences is used to bias data gathering in this forward simulation rollouts. It is worth noting that through Thompson sampling, as beliefs are propagated forward, the state evolution and observation functions f, g become deterministic maps from actions to hidden states and their corresponding observations.

At each time step *t*, the optimal action is chosen by solving an optimization problem for the expected look-ahead reward based on the predicted impact on the biased model evidence $r_t(\hat{y}_{t+1}) = \ln p(r(\hat{y}_{t+1})|\boldsymbol{\Phi}, u_t)$ of simulated data $\hat{y}_{t+1}$ for alternative stimuli in a compact set Ω:

$$u_t^{*} = argmax_{u \in \Omega}\, r_t(\hat{y}_{t+1}) \quad \textbf{(7)}$$

where r is a reward (or preference) function for observations and their underlying hidden states. The predicted observation $\hat{y}_{t+1}$ and the action u_t^{*} are then used to generate the posterior distributions for model parameters $q(\boldsymbol{\Phi})$ to be used at *t*+1. Thompson sampling is then applied to this simulated posterior and a new realization of model parameters is obtained and proceeds to next time step. The procedure is repeated until *t*=*H*. The main output from a belief propagation rollout is a sequence of optimized actions based on Thompson sampling of updated posteriors using simulation data over the planning sequence. As can be expected, the goodness of the se-

quence generated in a rollout depends on the value of the planning horizon *H*. Thus, to avoid significant sub-optimality losses only the initial part of the sequence is relevant. Moreover, as each rollout can be characterized by its cumulative expected reward for the actions in the sequence, it is advisable that for choosing the action that generates the most informative data, several rollouts must be simulated and then ranked them all based on the sequence of predicted rewards. Simulation of belief about beliefs rollouts based on Thompson sampling finesses the exploitation-exploration dilemma in relation to prior preferences in a modeling agent that predicts the free energy of the expect future for a given course of action over a receding-horizon *H*.

Inputs: H, current $\hat{x}_t$, prior $q(\boldsymbol{\Phi})$, state evolution and observation functions f, g

▷ **For** $t = 1$ to H

$\hat{q}(\boldsymbol{\Phi}_t) = q(\boldsymbol{\Phi})$ (Model parameter priors for rollout updates)

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

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

Simulate state transition using u_t^* and predict $r_{t+1}(\tilde{y_{t+1}^*}|u_t^*, \hat{x}_t)$

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

Infer next system state $\hat{x}_{t+1}$ for the optimal action u_t^*.

▷ **End for**

Output: rollout policy: $\pi^* = \{u_1^*, \ldots, u_H^*\}$ with its expexted cumulative reward $\mathcal{R}_k$

Fig. 3. Propagating beliefs about beliefs using posterior Thompson sampling.

5 Sophisticated active inference

In this section, we describe an efficient implementation of the proposed objective function for online redesign of planned interactions with the biological system in the context of reinforcement learning [14]. To select actions for purposely biasing the model, we iteratively optimise a policy π at each time step *t* using different samples from the distributions of model parameters [5]. A pseudocode for the proposed algorithm for using belief propagation in active inference is shown in Fig. 4. The internal loop has a forward rollout where the density $q(y_{t:T}, x_{t:T}, \Phi|\pi)$ is increasingly converted into a posterior density upon simulated data using a stagewise greedy redesign procedure based on Thompson sampling [12, 13] of the posterior distribution $q(\boldsymbol{\Phi})$ for model parameters. At each time step, an action is chosen by solving an optimization problem for the expected look-ahead reward based on the predicted impact on the biased model evidence $r_t(\hat{y}_{t+1}) = \ln\, p(r(\hat{y}_{t+1})|\boldsymbol{\Phi}, u_t)$ of simulated data $\hat{y}_{t+1}$ for alternative replanning decisions in a compact set Ω:

$$u_t^* = argmax_{u \in \Omega}\, r_t(\hat{y}_{t+1}) \quad \textbf{(8)}$$

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 $u_{t-1}^*, \hat{x}_{t-1}$

▷ **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 $r_{t+1}\left(\widehat{y_{t+1}^k} \middle| u_t^k, \hat{x}_t\right)$

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

End while

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

▷ **End for**

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

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

Interact with the system 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: $\pi^* = \{u_1^*, \ldots, u_{T-1}^*\}$, $\boldsymbol{y} = \{y_1, \ldots, y_T\}$, $\boldsymbol{r} = (r_1, \ldots, r_T)$, $q(\boldsymbol{\Phi})$

Fig. 4. Sophisticated active inference algorithm for model learning.

where r is a reward (or preference) function for observations and their underlying hidden states. The predicted observation $\hat{y}_{t+1}$ and the action u_t^* are then used to generate the posterior distribution for model parameters $\hat{q}(\boldsymbol{\Phi}_k)$ to be used at t+1. Thompson sampling is again applied and, using Eq. (4), the action u_{t+1}^* is calculated and the corresponding simulated observation $\hat{y}_{t+2}$ is computed. This forward rollout finishes when u_{T-1}^k is computed and the kth policy $\pi_k = \{u_t^k, \ldots, u_{T-1}^k\}$ is completely defined. For each iteration of the outer loop, a different policy π_k is computed. The generated sequences of planning actions are then ranked based on their corresponding cumula-

tive rewards $\mathcal{R}_k$ over the planning horizon $[t,..T]$. From the top ranked policy π_t^*, only the first action u_t^* is used for redesigning the sequence of future interactions and, at the next sampling time t+1 the observation y_{t+1} is obtained. Using experimental data (u_t^*, y_{t+1}), the joint posterior distribution $q(\Phi)$ for the parameters of the probabilistic model is updated using variational Bayesian inference (based on the Laplace approximation) and the master (external) loop begins re-estimating the optimal policy π_{t+1}^* from t+1 until the end of the planning horizon at time T.

6 Case study: Baker´s yeast production

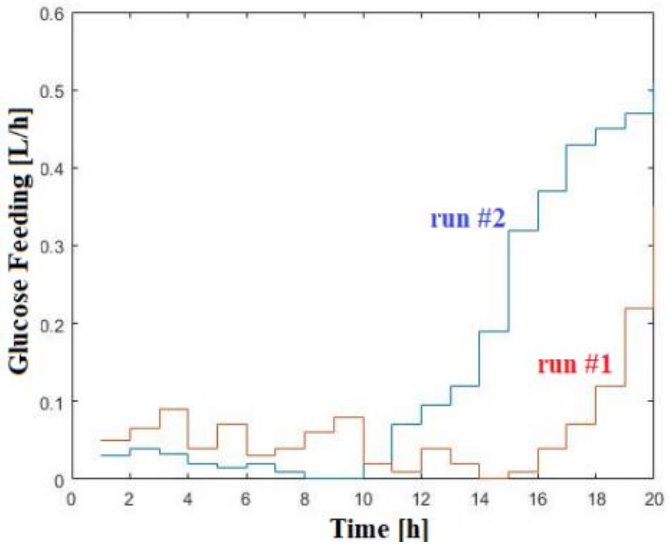
The macroscopic model proposed by Richelle et al. [11] for fed-batch baker's yeast production process in which the nitrogen and glucose consumptions are coordinated is used to test the sophisticated active inference algorithm in Fig. 4. The model includes a reaction in which the nitrogen and α-ketoglutarate are consumed to produce biomass. Also, the inhibition effect on glucose consumption by the accumulation of α-ketoglutarate is accounted for. The model has 15 parameters that define the reaction rates in biomass and ethanol production and consumption of glucose, nitrogen, and α-ketoglutarate. The concentration of α-ketoglutarate is a hidden state which is inferred using the model. The interested reader is referred to [11] for details.

The actions which are applied in each agent-system interaction are the feeding rate profiles for glucose and nitrogen. The duration of each modelling experiment is fixed to 20 hours, bioreactor initial conditions are known, and its content is sampled every hour to measure the concentration of biomass, ethanol, glucose, and nitrogen. Sample processing time is assumed equal to 30 min, leaving a maximum of 30 min to compute the redesign decision to the planned sequence of actions to be applied over the receding horizon. The parameter *K* in the algorithm (Fig. 4) is set to 10. It is worth noting that as the variance of parameter distributions is reduced the value of the hyper-parameter *K* can be reduced to advantage.

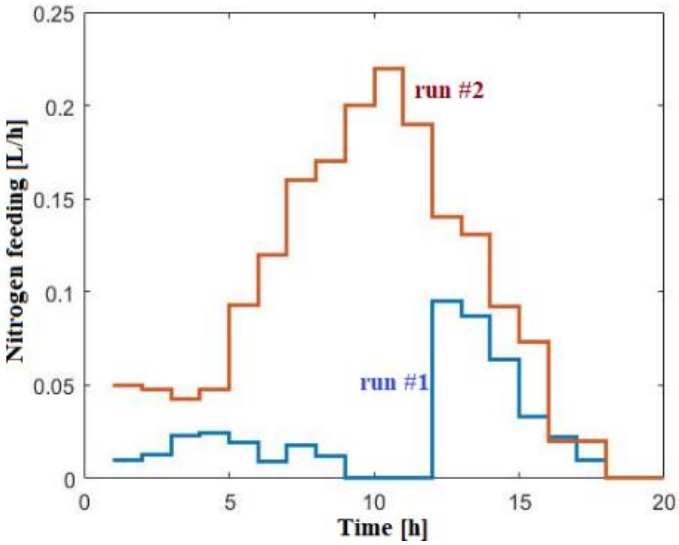
The aim is to purposefully bias model identification towards operating conditions that maximize the total amount of biomass that can be obtained at the end of the cultivation. Thus, the reward function r_k is defined to achieve a steady increase in the biomass concentration for consecutive samples. Results obtained are summarized in Fig. 5 and Fig. 6. Notice that due to initial uncertainty (prior distributions), for the first experiment exploration is significantly high. Also note that final biomass concentrations for modeling run #1 and # 2 in Table 3 are quite high despite they are not “one-shot” optimized feeding profiles using the probabilistic model with the updated parameter distributions. After two modeling experiments, as shown in Table 2, the parameter distributions have been updated to make the probabilistic model biased towards the most preferred region of system states [11], namely where biomass production is higher.

It is worth noting that the method bias data gathering in physiological states where nutrients are mainly used to produce biomass instead of ethanol which is a metabolite. Should production of ethanol (or expressing a given protein) were the objective this

will change the preference of the modeling agent. Accordingly, data gathering will be biased towards other physiological states.



(a) Glucose feeding rates



(b) Nitrogen feeding rates

Fig. 5. Substrate feeding profiles in run #1 and #2

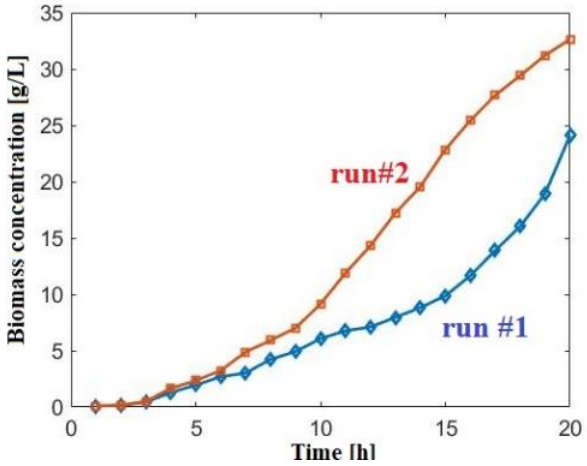


Fig. 6. Biomass production in run #1 and #2.

Table 1. Prior Normal distributions $N(\mu, \sigma^2)$ of the model parameters

		Mean	**Variance**
Parameter	*Units*	μ	σ
k_1	gX/gG	0.5431	0.250
k_2	gX/gG	0.0612	0.020
k_3	gX/gE	0.8929	0.085
k_4	gE/gG	0.2647	0.080
k_5	gA/gX/gG	0.2589	0.080
k_6	gX/gN	1.0150	0.150
μ_{Omax}	gG/gX/h	0.4445	0.125
μ_{Gmax}	gG/gX/h	2.5364	0.200
μ_{Nmax}	gN/gX/h	1.1903	0.150
K_G	gG/L	0.1524	0.030
K_I	gE/L	3.1817	0.050
K_N	gN/L	2.9370	0.050
K_A	gA/gX/L	9.0014	2.000
K_{IA}	gA/gX/L	5.5981	0.500
K_{IA2}	gA/gX/L	5.5737	0.210

7 Concluding remarks

A novel probabilistic method for modeling the dynamic behaviour of a biological system in the most preferred region of operating conditions is proposed. Based on simulation data and prior distributions, agent-system interaction sequences are redesigned online through active inference with belief propagation using Thompson sampling. Reinforcement learning is used to maximize the Bayesian model evidence, that is, to minimize surprise by implementing simulation rollouts when propagating beliefs about beliefs related to posterior distributions of model parameters once variational Bayesian updates are carried out.

An important advantage of the proposed approach is that it integrates the modeler´s preferences with goal-directed behaviour of biological systems from a cybernetic viewpoint using the free energy principle. In this setting, perception and action in biological systems minimize a free energy bound on Bayesian surprise. The free energy is thus an information-theoretic measure that bounds the current and the future expected statistical surprise, i.e., how unpredictable is a biological system under a given probabilistic model.

The free energy of the expected future in Eq. 6 quantifies using the Kullback-Leibler (KL) divergence (i.e., the distance) between the approximate and the true posterior distributions of system responses to external stimuli. According to the free energy principle, the modeling agent then acts in such a way as to minimize a free energy bound on the surprise at future time steps, i.e., Bayesian surprise which, informally speaking, provides a quantification of the difference between the agent's predictions about the system expected behavior to stimuli and the observed system responses.

Table 2. Posterior distributions $N(\mu, \sigma^2)$ for run #2 and run #3.

		Posterior Run #1		**Posterior** Run #2	
Parameter	*Units*	μ	σ	μ	σ
k_1	gX/gG	0.6232	0.112	0.7012	0.010
k_2	gX/gG	0.0578	0.011	0.0851	0.006
k_3	gX/gE	0.8500	0.005	0.8091	0.002
k_4	gE/gG	0.2450	0.009	0.2350	0.002
k_5	gA/gX/gG	0.0189	0.005	0.2162	0.037
k_6	gX/gN	0.9733	0.088	0.8817	0.044
μ_{Omax}	gG/gX/h	0.4210	0.039	0.4198	0.028
μ_{Gmax}	gG/gX/h	2.6472	0.145	2.7116	0.009
μ_{Nmax}	gN/gX/h	1.2279	0.067	1.2009	0.011
K_G	gG/L	0.1208	0.025	0.0989	0.010
K_I	gE/L	3.2011	0.443	2.9442	0.031
K_N	gN/L	2.9674	0.632	3.3598	0.278
K_A	gA/gX/L	9.4569	1.227	10.106	0.583
K_{IA}	gA/gX/L	5.8919	0.389	6.2991	0.267
K_{IA2}	gA/gX/L	6.131	0.201	5.8404	0.113

Table 3. Biomass final concentration in optimization runs.

Modeling Run #	Biomass [g/L]
1	24.11 (Final conc.)
2	32.64 (Final conc.)
Richelle et al. 2014 [11]	**32.00** (Exp. measurement)

Referencias

1. Blei, D. M, Kucukelbir A., McAuliffe, J. D.: Variational Inference: A Review for Statisticians, Journal of the American Statistical Association, 112:518 (2017) 859-877.

2. Daunizeau, J., Adam, V., Rigoux, L.: VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data, PLoS Comput Biol 10 (2014) 1, e1003441.

3. Daunizeau, J., Friston, K. J., Kiebel, S. J..: Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models, Physica D 238 (2009) 2089-2118.

4. de Jong H, Casagranda S, Giordano N, Cinquemani E, Ropers D, Geiselmann J, Gouze´ J-L. 2017 Mathematical modelling of microbes: metabolism, gene expression and growth. J. R. Soc. Interface 14 (2017)1-14 .20170502.

5. Friston, K., Da Costa, L., Hafner, D., Hesp, C., Parr, T.: Sophisticated inference, Neural Computation, 33 (2021) 713–763.

6. Kiparissides, A., Koutinas, M., Kontoravdi, C., Mantalaris, A., Pistikopoulos, E. N.: 'Closing the loop' in biological systems modeling — From the in silico to the in vitro, Automatica 47 (2011) 1147–1155.

7. Luna, M. F., Martínez, E. C.: Iterative modeling and optimization of biomass production using experimental feedback, Comput. Chem. Eng., 104 (2017) 151–163.

8. Martínez, E. C., Cristaldi, M. D., Grau, R. J.: Dynamic optimization of bioreactors using probabilistic tendency models and Bayesian active learning. Comput. Chem. Eng., 49 (2013) 37–49.

9. Parr, T., Friston, J. K.: Generalised free energy and active inference, Biological Cybernetics, 113 (2019) 495–513.

10. Ramkrishna, D., Song, H-S.: Cybernetic Modeling for Bioreaction Engineering. Cambridge University Press, Cambridge, New York (2019).

11. Richelle, A., Bogaerts, P.: Off-line optimization of baker's yeast production process, Chem. Eng. Sci., 119 (2014) 40–52.

12. Russo, D. J., Van Roy, B. (2014) Learning to Optimize via Posterior Sampling. Mathematics of Operations Research, 39 (2014) 1221-1243

13. Russo, D. J., Van Roy, B., Kazerouni, A., Osband, Ia., Wen, Z.: A Tutorial on Thompson Sampling. Foundations and Trends in Machine Learning, 11 (2018) 1-96.

14. Sutton, R. S., Barto, A. G.: Reinforcement learning: An introduction. 2nd Edn. MIT Press, Cambridge, MA (2018).

15. Tschantz, A., Seth A. K., Buckley, C. L.: Learning action-oriented models through active inference. PLoS Comput. Biol., 16 (2020) 4, e1007805.