Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 4608–4613

# Monitoring intracellular metabolite concentrations by moving horizon estimation based on kinetic modeling

**Sebastián Espinel-Ríos** * **Giulia Slaviero** * **Katja Bettenbrock** * **Steffen Klamt** * **Rolf Findeisen** **

* *Analysis and Redesign of Biological Networks, Max Planck Institute for Dynamics of Complex Technical Systems, Germany*

** *Control and Cyber-Physical Systems Laboratory, TU Darmstadt, Germany (e-mail: rolf.findeisen@iat.tu-darmstadt.de)*

**Abstract:** The biotechnology industry can significantly benefit from new paradigms such as smart manufacturing, digitalization and quality-by-design to render more competitive and robust processes. Real-time monitoring of key process parameters and performance indicators can facilitate the transition toward smart biomanufacturing. Since cells are typically used to catalyze biotechnological processes, online monitoring of the cell's health and intracellular metabolic status is of interest to the bioprocess industry. However, intracellular monitoring is challenging due to the intrinsic physical limitation imposed by the cell wall/membrane and the long time required for analytical measurements. This work outlines how soft sensors based on moving horizon estimation allow inferring the intracellular metabolite concentrations in bioprocesses, even for *advanced* or *complex* metabolic systems. For example, it can be applied for monitoring metabolic cybergenetic systems, whereby metabolic pathways are dynamically regulated via genetic circuits and external inputs. The moving horizon estimator uses a kinetic model of the central carbon and energy metabolism that describes the dynamics of the intracellular metabolites. We underline the use of moving horizon estimation considering the anaerobic fermentation of *Escherichia coli* with optogenetic regulation of the adenosine triphosphate turnover. With the information on the extracellular substrate, product and biomass concentrations, we could reconstruct the internal cell's metabolic state with a good performance.



*Keywords:* Bioprocess monitoring, moving horizon estimation, intracellular metabolites, kinetic modeling, fermentation.

## 1. INTRODUCTION

Biomanufacturing is currently undergoing a transition from trial-and-error and off-line approaches to quality-by-design and smart process control concepts. The ultimate goal is to maximize production (e.g., product yield, volumetric productivity, etc.) while achieving robust/consistent processes and high-quality products. Online monitoring of critical process parameters and key performance indicators, which are linked to the critical quality attributes of a product of interest, can lead to better process control. Therefore, moving toward smart biomanufacturing requires efficient process analytical technologies to enable real-time process monitoring (Gerzon et al., 2022).

There are already well-established sensors for measuring various *extracellular* process variables such as pH, temperature, osmolarity, partial pressure of gases, cell density and several substrates and products (Reyes et al., 2022). Comparatively, fewer studies focus on real-time monitoring of *intracellular* variables. In some cases, genetically encoded biosensors based on fluorescent proteins are used to measure the *in vivo* intracellular environment (cf. e.g., Torello Pianale et al. (2022)). However, such sensors are not straightforward to implement and use. First of all, even assuming orthogonality, they may impose a resource burden on the cell due to the intrinsic *cost* of synthesizing the biosensors' machinery (Mahr and Frunzke, 2016). Secondly, the output interpretation of fluorescence-based biosensors becomes cumbersome when the spectra of output signals overlap (Mahr and Frunzke, 2016).

Despite the advances in bioprocess monitoring, many important variables -especially at the intracellular level- cannot be measured or they can only be measured off-line. In this work, we focus on the use of soft sensors, i.e., sensors that rely on a mathematical model to infer unmeasured variables based on available process measurements (Luo et al., 2021; Reyes et al., 2022). In a previous study, we outlined a soft sensor for monitoring dynamic changes in *cell composition*, e.g., the intracellular distribution of catalytic enzymes, ribosomes and other elements such as non-catalytic proteins, RNA, DNA, storage compounds and lipids, assuming quasi-steady-state conditions of the intracellular metabolites (Espinel-Ríos et al., 2022b, 2023). Here, we describe a soft sensor for monitoring the dynamics of the *intracellular metabolites* during the process operation (online), i.e., the intermediate products of metabolism. Efficient and real-time monitoring of the intracellular metabolites can add significant value to the bioprocess monitoring toolbox since these species are linked to the resulting metabolic fluxes and productivity rates.

We exploit the concept of moving horizon estimation (MHE) to *reconstruct* the states of the system using a dynamic model of the process and the available past and present measurements. Some benefits of using MHE

⋆ This work was supported by the International Max Planck Research School for Advanced Methods in Process and Systems Engineering (IMPRS ProEng).


10.1016/j.ifacol.2023.10.968

include that it can handle non-linear equations, which is typically the case with biological models, and it can incorporate state and input constraints (Rawlings et al., 2020; Elsheikh et al., 2021). MHE has been previously proposed for parameter and state estimation of substrates and products in biotechnological processes (cf. e.g. Carius et al. (2018); Espinel-Ríos et al. (2022b, 2023); Tuveri et al. (2023)). In the previous references, the models constraining the MHE either lump up the intracellular metabolism or assume that the intracellular metabolites are in a quasi-steady state. Here we show that MHE can be applied to *advanced* or *complex* metabolic systems, e.g., metabolic cybergenetic systems, (Carrasco-López et al., 2020; Espinel-Ríos et al., 2023), for monitoring intracellular metabolite concentrations. In metabolic cybergenetics, the intracellular metabolic flux distribution is *dynamically* regulated via genetic circuits and external inputs, hence monitoring the *dynamic* profile of intracellular metabolites becomes pertinent for efficient quality and process control.

The remainder of this paper is structured as follows. We outline the MHE algorithm in Section 2. To capture the metabolic activity of cells, we use kinetic models developed using a *bottom-up* approach and parameter fitting as described in Section 3. Finally, in Section 4 we use the anaerobic glucose fermentation by *Escherichia coli* with optogenetic regulation of the adenosine triphosphate (ATP) turnover as a simulation example.

## 2. MOVING HORIZON ESTIMATION

Let us denote $x \in \mathbb{R}^{n_x}$ as the state vector, $u \in \mathbb{R}^{n_u}$ as the input vector, $\theta \in \mathbb{R}^{n_\theta}$ as the parameter vector, $y \in \mathbb{R}^{n_y}$ as the measurement vector and $w \in \mathbb{R}^{n_x}$ as the state noise vector. The system input and state constraints are defined by $c : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_c}$. Furthermore, let $F : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_x}$ be the system function and $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_y}$ be the measurement function. For simplicity, we consider a time-discrete version of the system model, with piece-wise constant input function over sampling times of size $\Delta t$. Also, let indicate with $(\cdot)_i$ a general optimization variable at time $t_i$. We will omit the explicit time-dependency of the variables when clear from the context. The MHE at time $t_k$ is formulated as an optimization problem with constraints (Rawlings et al., 2020; Elsheikh et al., 2021)

$$\min_{x_{k-N},\theta_k,w_{(\cdot)|k}} \quad J(x_{k-N},\theta_k,w_{(\cdot)|k}), \tag{1a}$$

$$\text{s.t.} \quad x_{i+1} = F(x_i,u_i,\theta_k) + w_i, \tag{1b}$$

$$y_i = h(x_i,u_i,\theta_k), \tag{1c}$$

$$0 \geq c(x_i,u_i,\theta_k), \tag{1d}$$

$$\text{for } i \in [k-N,k], \;\; k,N \in \mathbb{N}, \tag{1e}$$

where $N$ is the length of the estimation horizon, $x_{k-N}$ is the state vector at the beginning of the horizon and $w_{(\cdot)|k} := \{w_i, \forall i \in [k-N,k]\}$.

The objective function of the MHE optimization reads

$$J = \left\| \begin{bmatrix} x_{k-N} - \bar{x}_{k-N} \\ \theta_k - \bar{\theta} \end{bmatrix} \right\|_P^2 + \sum_{i=k-\bar{N}}^{k} \|\tilde{y}_i - y_i\|_R^2 + \|w_i\|_W^2, \tag{2}$$

where $\bar{x}_{k-N}$ and $\bar{\theta}$ are the *best guesses* of the states and the parameters at the beginning of the estimation horizon (e.g., previous estimates), $\tilde{y}_i$ are the measurements and $P$, $R$ and $W$ are weighting matrices. Note that $\|a\|_A^2 := a^\mathsf{T} A a$. The first term in Eq. (2) is also known as the *arrival cost*, which integrates the effect of the excluded past measurements into the estimation, i.e., the measurements laying further beyond the considered MHE horizon. The second term of Eq. (2) considers the difference between the measurements and the optimization variables, the *stage cost*, while the third term considers the state noise or the state *estimation error*. We assume that the parameters are constant over the considered horizon, however, if desired, they can be updated at each iteration of the MHE to account for uncertain parameters.

With the decision variables of the optimization problem described in (1), one can reconstruct the states via

$$F(x_i,u_i,\theta_k) = x_i + \int_{t_i}^{t_i+\Delta t} f(x,u,\theta_k)\mathrm{d}t, \tag{3}$$

where $f(x,u,\theta_k)$ is a vector-valued function that contains the dynamic equations of the system.

As already mentioned, the MHE uses a fixed estimation horizon, excluding the measurements located outside this *window*. When we take into account all the available measurements from the beginning of the process, then the horizon only grows in time, this special case of MHE is regarded as a full information estimation. However, one should be aware that the computational burden may increase with an increasing number of (past) measurements. Thus, in cases where a full information estimator is not computationally feasible, one can opt for an MHE approach with a shorter horizon. An overall scheme of the soft sensing strategy is presented in Fig. 1.

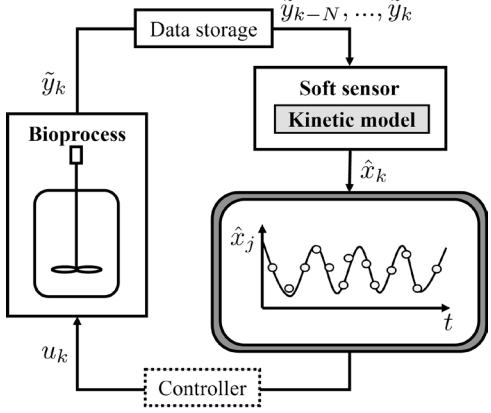



Fig. 1. Overall scheme of the soft sensing or state estimation strategy. The applied input can be obtained, e.g., via an appropriate controller. However, note that the scope of this work is limited to process monitoring.

## 3. KINETIC MODELING OF CELL METABOLISM

The best way to understand the phenotypic features of microorganisms and to track the time evolution of their metabolic components is through dynamic mathematical modeling. Kinetic models are usually set up using a *bottom-up* approach, i.e. *mechanistic* descriptions of the system's components are integrated to form a description of the system as a whole (Bruggeman and Westerhoff, 2007).

Following this approach, the first step of the modeling procedure requires knowledge of the network structure. This includes a description of how elements in the network are interconnected, e.g., information such as the stoichiometric matrix, enzyme regulation and mechanism, as well as the network compartments (Almquist et al., 2014).

The second step requires the identification of a suitable mathematical description for the reactions of the system.

Rate laws can be obtained from actual reaction mechanisms or can be approximated by general reaction kinetics (cf. Sauro (2012)). At this point, the model is written as a set of ordinary differential equations. Finally, the values for parameters and initial conditions are determined. These data can be collected from literature, databases and experiments. Usually, many parameter values are unknown and have to be determined through parameter estimation routines (a *top-down* approach).

For the deterministic case, a kinetic model has the general form

$$\frac{\mathrm{d}x}{\mathrm{d}t} = N \cdot \nu(x, u, \theta), \quad x(0) = x_0. \tag{4}$$

The vector $x$ usually consists of intracellular $z_{\text{in}} \in \mathbb{R}^{n_{z_{\text{in}}}}$ and extracellular $z_{\text{out}} \in \mathbb{R}^{n_{z_{\text{out}}}}$ metabolites, enzyme concentrations $E \in \mathbb{R}^{n_E}$, compartment volumes $v_c \in \mathbb{R}^{n_{v_c}}$, etc. Hence, $x := [z_{\text{in}}^\mathsf{T}, z_{\text{out}}^\mathsf{T}, E^\mathsf{T}, v_c^\mathsf{T}, ..]^\mathsf{T}$. The function of reaction rates is represented by $\nu : \mathbb{R}^{n_x} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\theta} \to \mathbb{R}^{n_\nu}$ and the matrix of stoichiometric coefficients is denoted as $N \in \mathbb{R}^{n_x \times n_\nu}$.

For the sake of generality, let us consider a simple reaction $i$, catalyzed by an enzyme $E_i$, i.e., $S \rightleftharpoons P$. Considering reversible Michaelis-Menten kinetics (Noor et al., 2013), the corresponding rate law $\nu_i$ can be written as

$$\nu_i = E_i \cdot k_{\text{cat}}^+ \frac{\frac{S}{k_S}\left(1 - \frac{k_{\text{cat}}^-}{k_{\text{cat}}^+}\frac{P/k_P}{S/k_S}\right)}{1 + \frac{S}{k_S} + \frac{P}{k_P}}, \tag{5}$$

where $k_{\text{cat}}^+$ and $k_{\text{cat}}^-$ are turnover constants of the forward and reverse reaction, respectively, while $k_S$ and $k_P$ are Michaelis-Menten constants.

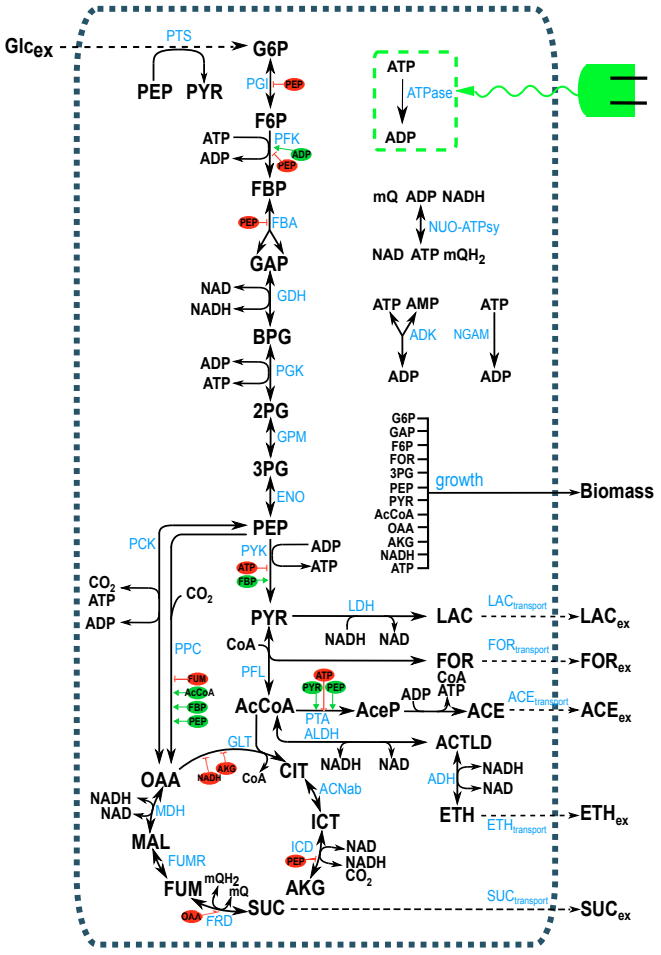
## 4. GLUCOSE FERMENTATION WITH OPTOGENETIC REGULATION

In Fig. 2, we present an overview of the *E. coli´*s kinetic model of the anaerobic central carbon and energy metabolism that is considered in this study. The model comprises 38 metabolites, 33 reactions and 253 parameters, and is adapted from the model published by Boecker et al. (2021). The adapted model is available upon request. Under anaerobic conditions, *E. coli* mainly synthesizes formate, ethanol and acetate. Enforcing the ATP turnover rate in the cell (i.e. introducing a mechanism that *wastes* an extra amount of energy) can lead to a higher glucose uptake rate and an increase in the yield of fermentation products (Boecker et al., 2019; Espinel-Ríos et al., 2022a).

The implementation of such ATP-wasting mechanism in the *conventional* way involves employing chemicals to induce the ATPase enzyme expression (Boecker et al., 2021). Note that the intracellular ATPase concentration is correlated with the ATPase-catalyzed reaction rate, i.e., the ATP-to-ADP hydrolysis rate. This strategy, however, requires a lot of empirical tuning and is very difficult to control. The use of optogenetics to modulate gene expression using light may be a more precise, fast and reliable solution (Hoffman et al., 2022). The online modulation of the ATP turnover via optogenetics has, in simulations, already proven to be a promising control strategy (Espinel-Ríos et al., 2022c, 2023). Therefore, here we assume that the intracellular ATPase enzyme concentration $z_{\text{ATPase}}$ can be fine-tuned using an external light input. This is unlike the kinetic model in Boecker et al. (2021) where $z_{\text{ATPase}}$ is considered constant for given induction levels by chemical inducers. Hence,

$$\frac{\mathrm{d}z_{\text{ATPase}}}{\mathrm{d}t} = \left(\alpha_1 + \alpha_2 \frac{I^{\alpha_3}}{k_e^{\alpha_3} + I^{\alpha_3}}\right)\mu - (d_e + \mu) z_{\text{ATPase}}, \tag{6}$$

where $I$ is the light intensity (hence, $u := I$), $\mu$ is the biomass growth rate, and $\alpha_1$, $\alpha_2$, $\alpha_3$, $k_e$ and $d_e$ are constant parameters. The parameter $\alpha_2$ represents the maximum light-dependant yield of ATPase per amount of biomass.




**Abbreviation of reaction names**: PTS: phoshotransferase system; PGI: glucose-6-phosphate isomerase; PFK: phosphofructokinase; FBA: fructose-bisphosphate aldolase (the associated reaction was lumped with the reaction of the triose-phosphate isomerase (TPI) thus yielding two molecules of GAP); GHD: glyceralde- hyde-3-phosphate dehydrogenase; PGK: phosphoglycerate kinase; ENO: enolase; GPM: phosphoglycerate mutase; PYK: pyruvate kinase; PFL: pyruvate formate lyase; LDH: lactate dehydrogenase; PTA: acetate kinase; ACK: phosphate acetyltransferase; ALDH: acetaldehyde-CoA dehydrogenase; ADH: alcohol dehydrogenase; PCK: phosphoenolpyruvate carboxykinase; PPC: phosphoenolpyruvate carboxylase; GLT: citrate synthase; ACNab: aconitate hydratase A and B; ICD: isocitrate dehydrogenase; MDH: malate dehydrogenase; FUMR: fumarase; FRD: lumped reaction of fumarate reductase; NUO-ATPsy: lumped reaction of ubiquinone oxidoreductase and atp synthesis; ADK: adenilate kinase; NGAM: consumption of ATP for non-growth associated maintenance; ATPase: reaction that hydrolyses ATP to ADP induced here from light. **Abbreviations of metabolites**: Glcex: external glucose (substrate); G6P: D-glucose-6-phosphate; F6P: D-fructose-6-phosphate; FBP: fructose-1,6-bisphosphate; GAP: D-glyceraldehyde-3-phosphate; BPG: 1,3-bisphospho-D-glycerate; 3PG: 3-phosphoglycerate; 2PG: 2-phosphoglycerate; PEP: phosphoenolpyruvate; PYR: pyruvate; AcCoA: acetyl coenzyme A; CoA: coenzyme A; CIT: citrate; ICT: iso-citrate; AKG: α-ketoglutarate; OAA: oxaloacetate; MAL: malate; FUM: fumarate; SUC: succinate; FOR: formate; LAC: lactate; ACE: acetate; ACTLD: acetaldehyde; ETH: ethanol; PRODUCTex: extracellular metabolites; ATP: adenosine triphosphate; ADP: adenosine diphosphate; AMP: adenosine monophosphate; NAD: lumped pool of oxidized nicotinamide adenine dinucleotide (NAD) and oxidized nicotinamide adenine dinucleotide phosphate (NADP); NADH: lumped pool of reduced nicotinamide adenine dinucleotide (NADH) and reduced nicotinamide adenine dinucleotide phosphate (NADPH); mQH2: menaquinol; mQ: menaquinone; CO2: carbon dioxide (fixed concentration).

Fig. 2. Scheme of the *E. coli* anaerobic metabolism used for the case study. Metabolites are shown in black; red and green circles represent inhibitors and activators, respectively. The green curly arrow represents the light-inducible expression of the ATPase enzyme. Refer to Boecker et al. (2021) for a complete description of the model structure.

The rate of the ATPase reaction $\nu_{\text{ATPase}}$ depends on the intracellular ATPase enzyme concentration as follows

$$\nu_{\text{ATPase}} = z_{\text{ATPase}} k_{\text{cat,ATPase}} \frac{z_{\text{ATP}}^\beta}{k_{\text{ATPase}}^\beta + z_{\text{ATP}}^\beta}, \tag{7}$$

where $z_{\text{ATP}}$ is the intracellular ATP concentration and $k_{\text{cat,ATPase}}$, $k_{\text{ATPase}}$ and $\beta$ are appropriate kinetic constants. The parameter values for Eqs. (6)-(7) are shown in Table 1.

Table 1. Nominal parameter values used in Eqs. (6)-(7).

| Parameter | Value | Unit | Ref. |
|---|---|---|---|
| $\alpha_1$ | $0.05\alpha_2$ | $\mu\text{mol/g}_\text{x}$ | See note 1* |
| $\alpha_2$ | 27.78 | $\mu\text{mol/g}_\text{x}$ | See note 1* |
| $\alpha_3$ | 2.49 | 1 | (Olson et al., 2014) |
| $k_e$ | 0.138 | $\text{W/m}^2$ | (Olson et al., 2014) |
| $d_e$ | $1.75 \times 10^{-5}$ | 1/s | (Benito et al., 1991) |
| $k_{\text{cat,ATPase}}$ | 1 | 1/s | See note 3* |
| $\beta$ | 3.752 | 1 | See note 3* |
| $k_{\text{ATPase}}$ | 0.78 | $\mu\text{mol/g}_\text{x}$ | See note 3* |

*Note 1. Assumed biologically reasonable value. The unit $\text{g}_\text{x}$ represents grams of biomass dry weight. *Note 2. The parameters $\alpha_3$ and $k_e$ correspond to a CcaS/CcaR optogenetic system, which is activated by green light and repressed by red light intensity. *Note 3. Adapted from Boecker et al. (2021).

The growth rate of the model, based on Monod kinetics, reads

$$\mu = \mu_{\max} \prod_{i=1}^{k} \frac{z_{\text{prc},i}}{k_{\text{prc},i} + z_{\text{prc},i}}, \tag{8}$$

where $\mu_{\max}$ is the maximum growth rate, $z_{\text{prc}} \in \mathbb{R}^{n_{z_{\text{prc}}}}$ is the vector of metabolite precursors of biomass formation (i.e., GAP, PEP, PYR, ATP, AcCoA, AKG, OAA, NADH, F6P, G6P, 3PG and FOR in the case study, cf. Fig. 2).

*Remark on simulation experiments.* The kinetic model described in this section, referred hereafter as the *nominal model*, is used to simulate the *real plant* and to generate the measurements for the estimation analysis in the next section.

### *4.1 Moving horizon estimation results*

To test our soft sensing strategy, we applied a predefined light input trajectory such that the cell metabolism is *excited* with different levels of ATPase gene expression throughout the process (see Fig. 3). We considered three different estimation scenarios:

- Scenario 1 (S1): all the extracellular metabolite (i.e., glucose, formate, lactate, ethanol, acetate and succinate) and biomass concentrations can be measured online. We assume no model-plant mismatch, i.e., the estimator uses the exact model of the plant to compute the MHE optimization.
- Scenario 2 (S2): same as S1, but with the addition of model-plant mismatch, obtained by disturbing two key nominal parameters of the model used by the estimator. Namely, both $\mu_{\max}$ and $\alpha_2$ were scaled-up by 5%. Note that the nominal value of $\alpha_2$ is listed in Table 1 and the nominal $\mu_{\max}$ was $5.9 \times 10^{-4}$ 1/s.
- Scenario 3 (S3): same as S2, but here a more pronounced model-plant mismatch is assumed, i.e., the nominal $\mu_{\max}$ and $\alpha_2$ were scaled-up by 10%. The disturbed model was used by the MHE to compute the optimization.

We added white Gaussian noise with 2.5% standard deviation to the *measurements*. We considered a horizon length $N = 7$. The weighting matrices were empirically fine-tuned such that $P = R = 1 \times 10^3 I_d$ and $W = 1 I_d$, where $I_d$ is the identity matrix of appropriate dimension. Note that we considered constant parameters over the entire process, thus they were not part of the estimation, nor updated at each iteration. For simplicity, we took equidistant sampling times, with piece-wise constant inputs. However, multi-rate estimation schemes are in principle also possible, whereby delayed measurements can be included in the estimation horizon once when they are available (Elsheikh et al., 2021). We used the HILO-MPC Python toolbox (Pohlodek et al., 2022) to solve all the estimation problems.

For all aforementioned scenarios, the estimated intracellular metabolite profiles are shown Fig. 3. Although strictly speaking the intracellular ATPase enzyme is not an intracellular *metabolite*, but rather an intracellular biomass *component*, it was also part of the estimation. To facilitate the comparison between the estimation scenarios, we computed the standard error (SE) of the estimate as

$$\text{SE} = \sqrt{\frac{\sum_{j=1}^{n_e} (z_{i,j} - \hat{z}_{i,j})^2}{n_e}}, \tag{9}$$

where $z_i$ is the *real* state value, $\hat{z}_i$ is the corresponding estimated value and $n_e$ is the number of estimates for $z_i$ throughout the process. The SE is a very intuitive figure since it has the same units as the estimated variable, thus it can be regarded as an *average error* of the soft sensor.

Before we show the estimation results, let us briefly remark that, as expected, the higher the light intensity the higher the specific production rate of the ATPase enzyme. This corresponds, in turn, to a decreasing ATP concentration. The ATPase intracellular concentration profile agrees in magnitude and trend with the input intensity; it is characterized by peaks, corresponding to the on-off cycles of the light.

In general, the MHE was able to predict well -to a greater or lesser extent- both qualitatively and quantitatively all concentration trends. In all cases, the estimates were in the same order of magnitude as the exact model values. Also, the fold changes of concentrations between a previous sampling point and the next one were respected in most of the cases. This means that all estimations were in good agreement with the nominal model, confirmed by the low SEs characterizing most species. The SE was slightly smaller or in the same range for S1 compared to S2. The same can be said for S2 in relation to S3. However, even for scenario S3, which contains the largest model-plant mismatch, the estimator still rendered a satisfactory performance, confirming the ability of the soft sensor to account for model uncertainty. We identified just a few metabolites where the estimator *struggled* to match the right value, although moving in the right direction (cf. e.g., the CIT, AcCoA, CoA and NAD and profiles).

Despite the good results, it might be that, depending on a specific model structure and the available measurements, not all the states can be *observed* by the estimator. In those cases, new measured states could be incorporated to satisfy observability. For instance, there are biosensors for measuring the ratio NADH/NAD (Liu et al. (2019)). Such a measurement could provide valuable information about the cellular redox state. Similarly, there are biosensors for measuring the intracellular ATP concentration (Deng et al. (2021)), which can be helpful for monitoring the energy dynamics in cells. Note that in the simulation example *observability* was *heuristically* checked by comparing the quality of the MHE estimate against the exact plant value via simulations. Although not performed here, one could in principle also *mathematically prove* observability by satisfying appropriate sufficient and/or necessary conditions (cf. e.g. Villaverde (2019)).

In summary, we believe that our soft sensor could be a powerful tool for bioprocess monitoring. Assuming we have

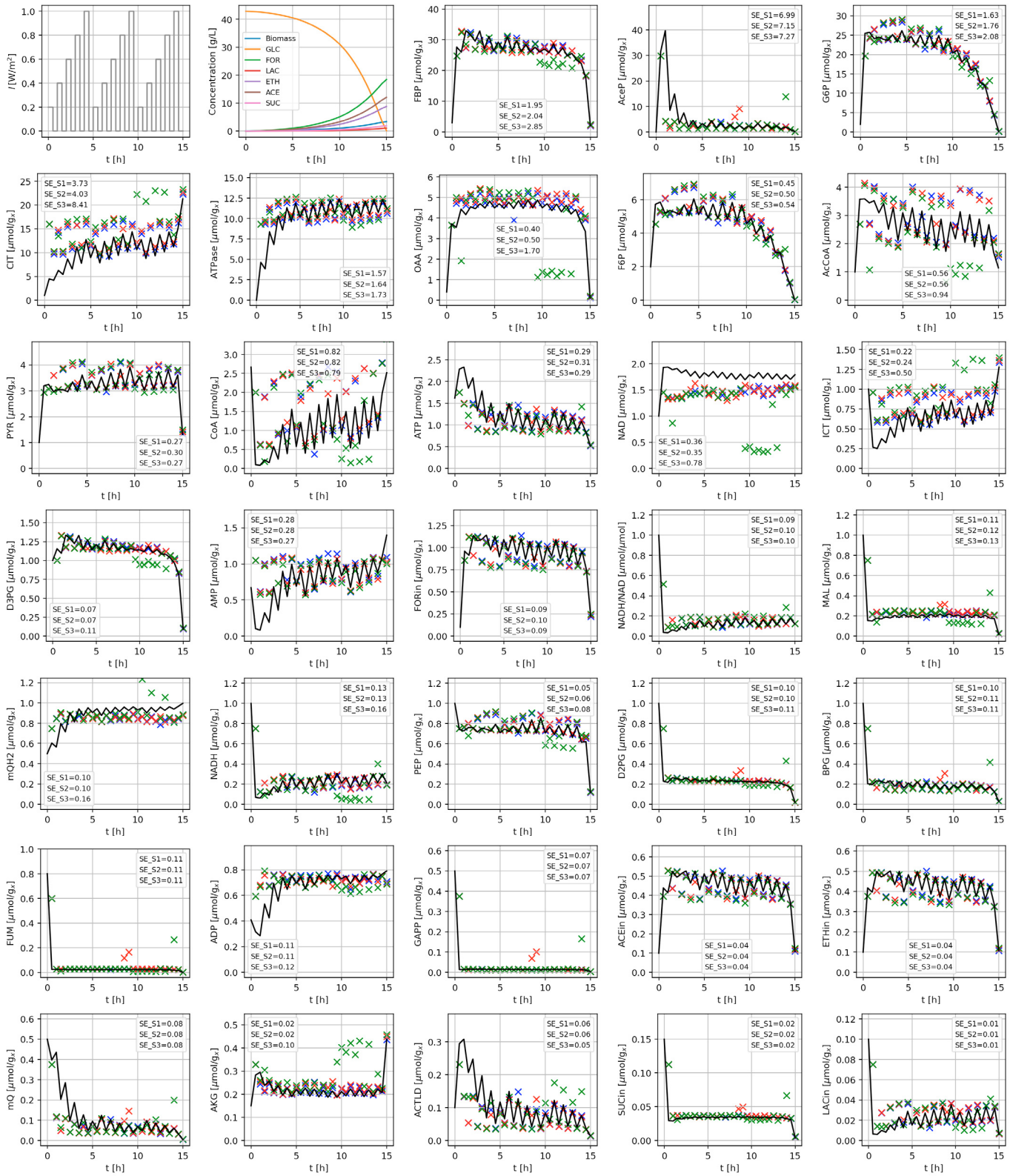



Fig. 3. The first two plots from left to right (top) show the dynamic input and extracellular concentrations. The remaining plots show different estimation scenarios of intracellular species with their corresponding standard error of the estimate. Notation for the estimation plots: exact value,—; estimation scenario 1 (S1), ×; estimation scenario 2 (S2), ×; estimation scenario 3 (S3), ×.

a sufficiently detailed and descriptive process model, such as the one used in the case study, we can infer relevant information on the intracellular bioprocesses based on limited available measurements. For instance, monitoring the accumulation of intracellular species can aid in the identification of metabolic bottlenecks and can suggest

down- or up-regulation of specific enzymes. Furthermore, we believe that our soft sensor may be a good candidate to support the use of model-based feedback control schemes such as model predictive control (Rawlings et al., 2020).

## 5. CONCLUSION AND OUTLOOK

In this work, we presented a strategy for monitoring the dynamic intracellular metabolite profile in bioprocesses by exploiting the concept of MHE. We outlined the use of a metabolic kinetic model that considers the dynamics of intracellular metabolites to constrain the estimation problem. As a simulation example, we used the anaerobic glucose fermentation by *E. coli* with optogenetic regulation of the ATPase enzyme for enforced ATP turnover applications. Even under model-plant mismatch and measurement noise, we were able to estimate the intracellular metabolites using only the measurements of biomass and extracellular metabolites which, with current technologies, can be in principle easily monitored online.

We believe that combining MHE approaches with detailed physiological kinetic models can provide a great tool for *virtual* process monitoring. Future work focuses on the experimental validation of the outlined soft sensor, as well as the exploration of different modeling strategies to constrain the estimation. For example, it would be possible to augment the first-principle biological models with data-driven parts to improve their prediction capacity, hence obtaining hybrid models. Furthermore, the development of surrogate models could potentially reduce the computational burden of the estimation. Finally, we are also interested in the integration of the described soft sensor into advanced feedback control schemes in the context of metabolic cybergenetic systems.

## REFERENCES


Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., and Jirstrand, M. (2014). Kinetic models in industrial biotechnology–improving cell factory performance. *Metab. Eng.*, 24, 38–60.

Benito, B., Moreno, E., and Lagunas, R. (1991). Half-life of the plasma membrane ATPase and its activating system in resting yeast cells. *Biochim. Biophys. Acta - Biomembr.*, 1063(2), 265–268.

Boecker, S., Slaviero, G., Schramm, T., Szymanski, W., Steuer, R., Link, H., and Klamt, S. (2021). Deciphering the physiological response of *Escherichia coli* under high ATP demand. *Mol. Syst. Biol.*, 17(12), e10504.

Boecker, S., Zahoor, A., Schramm, T., Link, H., and Klamt, S. (2019). Broadening the scope of enforced ATP wasting as a tool for metabolic engineering in *Escherichia coli*. *Biotechnol. J.*, 14(9), 1800438.

Bruggeman, F. and Westerhoff, H. (2007). The nature of systems biology. *Trends Microbiol.*, 15, 45–50.

Carius, L., Pohlodek, J., Morabito, B., Franz, A., Mangold, M., Findeisen, R., and Kienle, A. (2018). Model-based state estimation based on hybrid cybernetic models. *IFAC-PapersOnLine*, 51(18), 197–202.

Carrasco-López, C., García-Echauri, S.A., Kichuk, T., and Avalos, J.L. (2020). Optogenetics and biosensors set the stage for metabolic cybergenetics. *Curr. Opin. Biotechnol.*, 65, 296–309.

Deng, Y., Beahm, D.R., Ionov, S., and Sarpeshkar, R. (2021). Measuring and modeling energy and power consumption in living microbial cells with a synthetic ATP reporter. *BMC Biol.*, 19(1), 101.

Elsheikh, M., Hille, R., Tatulea-Codrean, A., and Krämer, S. (2021). A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. *Comput. Chem. Eng.*, 146, 107219.

Espinel-Ríos, S., Bettenbrock, K., Klamt, S., and Findeisen, R. (2022a). Maximizing batch fermentation efficiency by constrained model-based optimization and predictive control of adenosine triphosphate turnover. *AIChE Journal*, 68(4), e17555.

Espinel-Ríos, S., Morabito, B., Bettenbrock, K., Klamt, S., and Findeisen, R. (2022b). Soft sensor for monitoring dynamic changes in cell composition. *IFAC-PapersOnLine*, 55(23), 98–103.

Espinel-Ríos, S., Morabito, B., Pohlodek, J., Bettenbrock, K., Klamt, S., and Findeisen, R. (2022c). Optimal control and dynamic modulation of the ATPase gene expression for enforced ATP wasting in batch fermentations. *IFAC-PapersOnLine*, 55(7), 174–180.

Espinel-Ríos, S., Morabito, B., Pohlodek, J., Bettenbrock, K., Klamt, S., and Findeisen, R. (2023). Towards a modeling, optimization and predictive control framework for fed-batch metabolic cybergenetics. arXiv:2302.02177.

Gerzon, G., Sheng, Y., and Kirkitadze, M. (2022). Process analytical technologies – advances in bioprocess integration and future perspectives. *J. Pharm. Biomed. Anal.*, 207, 114379.

Hoffman, S.M., Tang, A.Y., and Avalos, J.L. (2022). Optogenetics illuminates applications in microbial engineering. *Annu. Rev. Chem. Biomol. Eng.*, 13(1), 373–403.

Liu, Y., Landick, R., and Raman, S. (2019). A regulatory NADH/NAD+ redox biosensor for bacteria. *ACS Synth. Biol.*, 8(2), 264–273.

Luo, Y., Kurian, V., and Ogunnaike, B.A. (2021). Bioprocess systems analysis, modeling, estimation, and control. *Curr. Opin. Chem. Eng.*, 33, 100705.

Mahr, R. and Frunzke, J. (2016). Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol.*, 100(1), 79–90.

Noor, E., Flamholz, A., Liebermeister, W., Bar-Even, A., and Milo, R. (2013). A note on the kinetics of enzyme action: A decomposition that highlights thermodynamic effects. *FEBS Lett.*, 587(17), 2772–2777.

Olson, E.J., Hartsough, L.A., Landry, B.P., Shroff, R., and Tabor, J.J. (2014). Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods*, 11(4), 449–455.

Pohlodek, J., Morabito, B., Schlauch, C., Zometa, P., and Findeisen, R. (2022). Flexible development and evaluation of machine-learning-supported optimal control and estimation methods via HILO-MPC. arXiv:2203.13671.

Rawlings, J., Mayne, D., and Diehl, M. (2020). *Model predictive control: theory, computation and design*. Nob Hill Publishing, LLC, Santa Barbara, 2nd edition.

Reyes, S.J., Durocher, Y., Pham, P.L., and Henry, O. (2022). Modern sensor tools and techniques for monitoring, controlling, and improving cell culture processes. *Processes*, 10(2), 189.

Sauro, H.M. (2012). *Enzyme kinetics for systems biology*. Ambrosius Publishing, Lexington, KY, 1st edition.

Torello Pianale, L., Rugbjerg, P., and Olsson, L. (2022). Real-time monitoring of the yeast intracellular state during bioprocesses with a toolbox of biosensors. *Front. Microbiol.*, 12, 802169.

Tuveri, A., Nakama, C.S., Matias, J., Holck, H.E., Jäschke, J., Imsland, L., and Bar, N. (2023). A regularized moving horizon estimator for combined state and parameter estimation in a bioprocess experimental application. *Comput. Chem. Eng.*, 172, 108183.

Villaverde, A.F. (2019). Observability and structural identifiability of nonlinear biological systems. *Complexity*, 2019, 1–12.