Computers and Chemical Engineering 172 (2023) 108183

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

A regularized Moving Horizon Estimator for combined state and parameter estimation in a bioprocess experimental application

Andrea Tuveri [a], Caroline S.M. Nakama [a], José Matias [a], Haakon Eng Holck [a], Johannes Jäschke [a], Lars Imsland [b], Nadav Bar [a,*]

[a] Department of Chemical Engineering, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[b] Department of Engineering Cybernetichs, Norwegian University of Science and Technology (NTNU), Trondheim, Norway

ARTICLE INFO

Keywords:
State estimation
Parameter estimation
Moving Horizon Estimator
Regularization
Optimization
Bioprocess

ABSTRACT

Due to the lack or high costs of measurement devices to monitor and control metabolites in microbial cultivation processes, state estimators are often required. These estimators depend on available *on-line* measurements and model dynamics. However, they are often characterized by simple models due to the lack of full knowledge on the process dynamics and high variability in the cell metabolism. This causes uncertainty in the model parameters and therefore the necessity of *on-line* model adaptation, for instance through simultaneous state and parameter estimation. However, these estimation problems are often ill conditioned. The Moving Horizon Estimator (MHE) is a good candidate in this context, since it easily allows enforcing hard constraints as well as regularization to address the ill-posedness. In this work, we present a method for simultaneous state and parameter estimation in the absence of full state measurements, with the aid of two regularization methods, in a microbial fed-batch cultivation.

1. Introduction

Real-time monitoring is of paramount importance in control of bioprocesses. Indeed, it is critical for quality assessment and therefore feedback implementations. However, in many applications this is a great challenge, since not all the variables of interest can be directly measured *on-line*. In the case that key variables are unmeasured, state estimators can be used to infer information from the available measurements to compensate for it (Doyle III, 1998; Dochain, 2003). State estimators, also called Soft Sensors (Dochain, 2003), depending on mathematical models (Rao, 2000) that usually represent the real system in an oversimplified manner (e.g. Monod growth model, Monod (1949)), can yield to poor estimates, due to the high model uncertainty, mostly caused by biological variability (Jabarivelisdeh et al., 2020).

Schei (2008) and Mohd Ali et al. (2015) reviewed the implementation of state estimators in chemical processes, discussing design issues, as well as general guidelines for selecting the proper technique in specific applications. Among the various state estimators reported, optimization-based methods, such as the Moving Horizon Estimator (MHE), have two main advantages: (1) they can handle multi-rate measurements easily (Rao, 2000), often encountered in bioprocesses (Elsheikh et al., 2021); (2) they can explicitly incorporate equality or inequality constraints in both states and parameters that may vary in the model (Rao, 2000; Rao et al., 2003), enabling the possibility to handle uncertainty (Doyle III, 1998; Dochain, 2003).

Whereas the capability of handling constraints in combination with multi-rate measurements has been extensively exploited for biological applications, to the best of our knowledge, constraints handling methods have been mostly applied to either state or parameter estimation, not both. Our work aims to account for model uncertainty by *simultaneously* performing state and parameter estimation in absence of full state feedback. Although not applied to a biological process, the work of Liu et al. (2021) also recently presented simultaneous state an parameter estimation in an MHE. However, differently from the work of Bae et al. (2021) and ours, they used a penalty quadratic term for the calculation of the arrival cost (i.e. to take the past information into account).

The majority of the up to date MHE applications in bioprocesses, reconstruct both states and key parameters using multi-rate full state feedback (Gatzke and Doyle III, 2002; Küpper et al., 2009, 2010; Vercammen et al., 2016; Jabarivelisdeh et al., 2020; Elsheikh et al., 2021; Bae et al., 2021; Hernández Rodríguez et al., 2021; Valipour and Ricardez-Sandoval, 2021, 2022b,a). Some of the reported applications (Gatzke and Doyle III, 2002; Küpper et al., 2009, 2010; Vercammen et al., 2016; Hernández Rodríguez et al., 2021) have shown successful reconstruction of critical model parameters (i.e. model update or adaptation). Elsheikh et al. (2021) presented a review on the

* Corresponding author.
E-mail address: nadi.bar@ntnu.no (N. Bar).

https://doi.org/10.1016/j.compchemeng.2023.108183
Received 1 July 2022; Received in revised form 4 February 2023; Accepted 15 February 2023
Available online 17 February 2023

utilization of multi-rate measurements and the effect of different arrival cost updates. Valipour and Ricardez-Sandoval (2021) and Jabarivelisdeh et al. (2020) showed the implementation of an MHE along with a nonlinear Model Predictive Controller (NMPC), while Valipour and Ricardez-Sandoval (2022b,a) focused on the improvement of the estimation of unexpected process and measurement noises, by the adoption of non-Gaussian distributions. Bae et al. (2021) presented instead an implementation that attenuates the ill-posed problem arising from the estimation of all parameters, by reducing the number of decision variables (i.e. regularization method). Fewer and less recent works focused on inferring information on the unmeasured states (Raïssi et al., 2005; Goffaux and Wouwer, 2008) without taking into account the possibility to estimate the model parameters. The work of Raïssi et al. (2005) presented an implementation where the solution of the optimization is a bounded set consistent with model, measurements and errors. Goffaux and Wouwer (2008), on the other hand, implemented a min–max optimization problem to deal with parameter uncertainty.

The concept of expressing the uncertain parameters as bounded variables can be dated back to the work of Grossmann and Sargent (1978) and later addressed by Bonvin et al. (2001). Considering that MHE has the ability to explicitly incorporate physical constraints on the parameters, and that disturbances in the form of time-varying parameters can be added as extra degrees of freedom in the optimization problem (Robertson et al., 1996; Rao et al., 2003; Kühl et al., 2011), MHE results a good candidate for performing combined state and parameter estimation. However, uncertainty in the model parameters can potentially cause a large bias in the estimates of the unmeasured states (Dochain, 2003). Moreover, the inclusion of process parameters as decision variables, can lead to an ill-posed optimization problem. Therefore, to obtain a unique local solution, regularization methods are required.

Several approaches have been proposed for selecting parameters subsets (i.e. regularization methods) to attenuate ill-posed parameter estimation problems. McLean and McAuley (2012) reported and divided them in different categories: (1) methods based on the correlation and collinearity indexes (López C et al., 2015; Anane et al., 2019), (2) Orthogonalization Method (Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009; Bae et al., 2021), (3) methods based on the Fisher Information Matrix (FIM) characteristics (Balsa-Canto et al., 2007), and (4) methods based on Principal Component Analysis (PCA) and eigenvalue–eigenvector decomposition (Vajda et al., 1989; Kim and Lee, 2019; Nakama et al., 2020; Chen et al., 2022). In addition to these methods, common practice is also to allow the variation of only one poorly known parameter per measurement value (Dochain, 2003). Further details on regularization techniques can also be found in Kravaris et al. (2013). Most of the regularization methods are dependent on cut-off or threshold values for the selection of the number of constraints. However, some authors already presented alternatives, to avoid cut-off values, based on the minimization of the mean squared error (MSE) to improve model prediction (Chu et al., 2009; Wu et al., 2011).

To capture the changing or missing dynamics caused by the plant-model mismatch (Psichogios and Ungar, 1992; Jabarivelisdeh et al., 2020; Bae et al., 2021), our work exploits the ability of MHE to explicitly incorporate physical constraints on the parameters, to continuously update them.

This is done by adding them as single degrees of freedom to the optimization problem and, as also reported in Bae et al. (2021), will result in parameter drifts. To overcome the ill-conditionedness of the problem that arises from that, we implemented and compared two different regularization methods, coupled with a stopping criterion based on structural identifiability. The latter allowed to select, at each time point, the number of active constraints necessary to regularize the problem, based on the available *on-line* information. Firstly, to reduce the search space of the decision variables, we adopted the regularization approach (Subset Selection by Transformation, SST) as proposed by Kim and Lee (2019) and implemented it in an MHE. This was done using real experimental data of a fed-batch bacterial cultivation of *Corynebacterium glutamicum* (Tuveri et al., 2021). Secondly, to evaluate the properties of SST and show its advantages, we compared it to the known Orthogonalization Method (OM) (Lund and Foss, 2008; Bae et al., 2021). Finally, to avoid using a stopping criterion based on a threshold value, for the selection of the number of the active constraints on the parameters, we proposed a strategy based on the structural identifiability of the system (Villaverde, 2019), in both OM and SST. The results, validated using experimental data from a real bioprocess, present accurate state estimates with respect to the reference values also allowing simultaneous parameter adaptation.

2. Experimental setup

The experiment was conducted using the *C. glutamicum* ATCC13032 strain. The cells were harvested in a shake flask pre-culture over night in 2YT complex medium and then inoculated in a 2.7 L baffled stirred tank reactor Labfors5 (Infors AG, Switzerland). The experiment was conducted with initial volume of 1.5 L and initial OD_{600} of 1. The 500 mL of glucose feeding (100 g/L) were added once the dissolved oxygen stabilized above 60%. Temperature, pressure and pH were kept respectively to 30 °C, 1 bar and 7. The reactor was aerated with 2 NL/min pressurized air and the dissolved oxygen was controlled above 30% by modifying the stirrer speed (200–1100 rpm). *C. glutamicum* was cultured on CGXII minimal medium and glucose used as carbon source. *On-line* measurements for absorbance (840–910 nm, ASD12-N Absorption Probe, Optek GmbH), volume and *off-gas* composition (BlueInOne Ferm, BlueSens GmbH) were available every 10 s trough the process information management system Lucullus (Securecell, Switzerland). *Off-line* samples for sugars were instead collected and stored at 4 °C by the NUMERA system (Securecell, Switzerland). Glucose samples were analyzed using an high-pressure liquid chromatography system (UltiMate 3000 series, Thermo Scientific, U.S.). The *off-line* biomass was instead evaluated by measuring cell dry weight. The reader is referred to Tuveri et al. (2021) for more details on cultivation process and analytical procedures.

3. Bioprocess description

This section starts by introducing the system dynamics (Section 3.1) and the noise structure (Section 3.1.1). Following this, *on-line* (i.e. outputs) and *off-line* measurements (i.e. reference values) are described (Section 3.1.2) together with the signal processing (Section 3.1.3).

3.1. System model

The process describes an aerobic bacterial cultivation, which is performed in two phases:

- A batch phase, up to complete sugar depletion and stabilization of dissolved oxygen above 60%;
- A fed-batch phase, followed a short period of starvation.

The system dynamics are described using first order Monod-like kinetics (Tuveri et al., 2021). In addition, cell death is considered to be linear and dilution is due to the addition of feeding. The model equations are the following:

$$\begin{aligned}
\dot{V} &= F_{in} \\
\dot{X} &= -\frac{F_{in}}{V} X + \mu_{max} \frac{S}{K_s + S} X - k_d X \\
\dot{S} &= \frac{F_{in}}{V}(S_{in} - S) - \mu_{max} \frac{S}{K_s + S} \frac{X}{Y_{XS}} \\
\dot{CO}_2 &= \mu_{max} \frac{S}{K_s + S} \frac{X}{Y_{XCO_2}} - q_{air} CO_2
\end{aligned} \tag{1}$$

The states V, X, S and CO_2 in Eq. (1) represent the concentrations of volume, biomass, glucose (or substrate, sugar) and carbon dioxide

Table 1
Model parameters $\theta \in \mathbb{R}^{n_\theta}$ for the system in Eq. (1) given values, units and standard deviation (Tuveri et al., 2021).

Parameter	Description	Value	Unit	Std. Dev.
μ_{max}	Maximum growth rate	0.19445	$[h^{-1}]$	$3.25 \cdot 10^{-6}$
K_S	Monod constant	0.007	$[g\ L^{-1}]$	$3.92 \cdot 10^{-6}$
k_d	Death rate constant	0.006	$[h^{-1}]$	$4.49 \cdot 10^{-6}$
Y_{XS}	S from X yield	0.42042	$[g\ g^{-1}]$	$3.58 \cdot 10^{-6}$
$Y_X CO_2$	CO_2 from X yield	0.54308	$[g\ g^{-1}]$	$2.22 \cdot 10^{-6}$

(CO_2), respectively. F_{in} indicates the feeding flow to the reactor, given a substrate concentration S_{in}. In addition, q_{air} represents a constant gas inflow. From herein, we define:

- State vector $x \in \mathbb{R}^{n_x}$ and input $u \in \mathbb{R}^{n_u}$:

$$x = [V,\ X,\ S,\ CO_2]^T \quad \text{and} \quad u = [F_{in}] \tag{2}$$

- Measured outputs $y \in \mathbb{R}^{n_y}$:

$$y = [V,\ X,\ CO_2]^T \tag{3}$$

- Parameters $\theta \in \mathbb{R}^{n_\theta}$:

$$\theta = [\mu_{max}, K_s, k_d, Y_{XS}, Y_{XCO_2}]^T \tag{4}$$

This allows us to write the model in Eq. (1) as $\dot{x} = f(x, u, \theta)$. Additionally, by adding the parameters as state variables with zero dynamics ($\dot{\theta}_i = 0$), the system in Eq. (1) can be written as:

$$\dot{x}_a = \bar{f}(x_a, u) \tag{5}$$

where $x_a \in \mathbb{R}^{n_x+n_\theta}$ is defined as $x_a = [x, \theta]$.

3.1.1. Noise structure tuning

Following the works of Leu and Baratti (2000) and Kolås et al. (2009), the process noise was added to the deterministic plant model (Eq. (1)) twofold. In the states, in order to address unmodeled dynamics, and in the parameters, to address parameter uncertainty (standard deviations are reported in Table 1). Indeed, the idea is that the model uncertainty (Leu and Baratti, 2000; Kolås et al., 2009) arises mostly from the uncertainty in the model parameters ($\omega_{\mu_{max}}, \omega_{K_S}, \omega_{k_d}, \omega_{Y_{XS}}, \omega_{Y_{XCO_2}}$). Moreover, in our case, to prevent the noise covariance matrix from becoming singular when the substrate S is depleted, the noise is added to the state dynamics ($\omega_V, \omega_X, \omega_S, \omega_{CO_2}$). The noise enters the material balances (Eq. (1)) as:

$$\begin{aligned}
\dot{V} &= F_{in} + \omega_V \\
\dot{X} &= -\frac{F_{in}}{V} X + (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{K_S}) + S} X - (k_d + \omega_{k_d}) X + \omega_X \\
\dot{S} &= \frac{F_{in}}{V}(S_{in} - S) - (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{K_S}) + S} \frac{X}{(Y_{XS} + \omega_{Y_{XS}})} + \omega_S \\
\dot{CO}_2 &= (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{K_S}) + S} \frac{X}{(Y_{XCO_2} + \omega_{Y_{XCO_2}})} - q_{air} CO_2 + \omega_{CO_2}
\end{aligned} \tag{6}$$

given the noise vector $\omega \in \mathbb{R}^{(n_x+n_\theta)}$:

$$\omega = \left[\omega_{\mu_{max}}\ \omega_{k_s}\ \omega_{k_d}\ \omega_{Y_{XS}}\ \omega_{Y_{XCO_2}}\ \omega_V\ \omega_X\ \omega_S\ \omega_{CO_2} \right]^T$$

with $\omega \sim \mathcal{N}(0, Q_\omega)$, where the diagonal elements of the covariance matrix $Q_\omega \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$ are reported in Table 2. For the batch phase, the values of Q_ω related to the parameters θ are tuned by setting them equal to the value of the variance, obtained by parameter estimation on a single experiment, while the ones related to the states x are defined as tuning parameters and therefore selected manually. For the fed-batch phase, only the values of $\sigma^2_{K_S}$ and $\sigma^2_{Y_{XCO_2}}$ are increased to compensate for unmodeled dynamics (Tuveri et al., 2021). The model with noise (Eq. (6)) can be compactly written as:

$$\dot{x} = f(x, u, \theta, \omega) \tag{7}$$

Table 2
Diagonal elements of the covariance matrix Q_ω. The values are kept constant until the fed-batch phase, when the values of $\sigma^2_{K_S}$ and $\sigma^2_{Y_{XCO_2}}$ are increased to compensate for unmodeled dynamics.

Variance	Additive noise	Batch	Fed-batch	Unit
$\sigma^2_{\mu_{max}}$	In μ_{max}	$1.05 \cdot 10^{-11}$	–	$[h^{-2}]$
$\sigma^2_{K_S}$	In K_S	$1.54 \cdot 10^{-11}$	$3.38 \cdot 10^{-2}$	$[g^2\ L^{-2}]$
$\sigma^2_{k_d}$	In k_d	$2.02 \cdot 10^{-11}$	–	$[h^{-2}]$
$\sigma^2_{Y_{XS}}$	In Y_{XS}	$1.28 \cdot 10^{-11}$	–	$[g^2\ g^{-2}]$
$\sigma^2_{Y_X CO_2}$	In $Y_X CO_2$	$4.91 \cdot 10^{-12}$	$4.91 \cdot 10^{-2}$	$[g^2\ g^{-2}]$
σ^2_V	In V	$1 \cdot 10^{-1}$	–	$[L^2\ h^{-2}]$
σ^2_X	In X	$1 \cdot 10^{-2}$	–	$[g^2\ h^{-2}]$
σ^2_S	In S	$1 \cdot 10^{-2}$	–	$[g^2\ h^{-2}]$
$\sigma^2_{CO_2}$	In CO_2	$1 \cdot 10^{-1}$	–	$[h^{-2}]$

The process noise covariance $Q_k \in \mathbb{R}^{n_x \times n_x}$ for the system described as in Eq. (6), is updated at each sampling time k as Tuveri et al. (2021):

$$Q_k = G_k \cdot Q_\omega \cdot {G_k}^T \tag{8}$$

Eq. (8) therefore allows to have a state-dependant varying covariance Q_k, where Q_ω is a constant related to the statistics of the parameter uncertainty. $G_k \in \mathbb{R}^{n_x \times (n_\theta + n_x)}$ is the Jacobian of Eq. (7) with respect to the noise ω:

$$G_k = \frac{\partial f(x, u, \theta, \omega)}{\partial \omega} \tag{9}$$

Additionally, to estimate the parameters θ together with the states x, we define them as additional state variables and the augmented state vector becomes $x_a \in \mathbb{R}^{n_x+n_\theta}$. In accordance with the work of Grossmann and Sargent (1978), the parameters are considered as bounded variables, given that probability distribution functions for the parameters are available. This is done to reflect the drifting characteristics of the model parameters used to describe the system, as was previously showed in Bae et al. (2021). Therefore, by approximating the probability distribution functions of the parameters to be normal, with mean value θ_0 and variance ${\sigma_\theta}^2$ (Table 2), we will treat them as drifting bounded variables, allowing them to vary within their bounds at each iteration. The augmented model (Eq. (10)) can therefore be written as:

$$\begin{aligned}
\dot{V} &= F_{in} + \omega_V \\
\dot{X} &= -\frac{F_{in}}{V} X + (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{K_S}) + S} X - (k_d + \omega_{k_d}) X + \omega_X \\
\dot{S} &= \frac{F_{in}}{V}(S_{in} - S) - (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{K_S}) + S} \frac{X}{(Y_{XS} + \omega_{Y_{XS}})} + \omega_S \\
\dot{CO}_2 &= (\mu_{max} + \omega_{\mu_{max}}) \frac{S}{(K_s + \omega_{K_S}) + S} \frac{X}{(Y_{XCO_2} + \omega_{Y_{XCO_2}})} - q_{air} CO_2 + \omega_{CO_2} \\
\dot{\mu}_{max} &= \omega_{\mu_{max}} \\
\dot{K}_s &= \omega_{K_s} \\
\dot{k}_d &= \omega_{k_d} \\
\dot{Y}_{XS} &= \omega_{Y_{XS}} \\
\dot{Y}_{XCO_2} &= \omega_{Y_{XCO_2}}
\end{aligned} \tag{10}$$

Eq. (10) can be written compactly as $\dot{x}_a = \bar{f}(x_a, u, \omega)$. The state-dependant covariance matrix for the augmented system $\bar{Q}_k \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$ is updated at each sampling time k as:

$$\bar{Q}_k = \begin{bmatrix} Q_k & 0 \\ 0 & Q_\theta \end{bmatrix} \tag{11}$$

where the state-dependant submatrix Q_k (Eq. (8)) is updated at each sampling time k, while $Q_\theta \in \mathbb{R}^{n_\theta \times n_\theta}$ is a constant diagonal matrix with variances ${\sigma_\theta}^2$ defined in Table 2. The latter is done to allow the parameters to vary within the bounds at each iteration k and it is a

necessary assumption since the real dynamics of the parameters are unknown.

3.1.2. On-line and off-line measurements

The *on-line* output measurements y were available with a frequency of 10 s and used by the estimator to monitor biomass, volume and CO_2 and to infer the unmeasured glucose composition. Additionally, to have reference values for evaluating the estimation performance, *off-line* samples of X and S were taken with a lower sample frequency and not used at any time by the estimator. The samples for biomass X were collected manually approximately every 3 h, whereas for glucose S they were collected automatically by the NUMERA auto-sampler every hour. Additional information is available in Tuveri et al. (2021).

3.1.3. Signal processing

The *on-line* output measurements y were used by the estimator every 60 s. Signals from the absorbance probe were obtained in concentration units (0.05–4 CU) and than converted to g/L of biomass X (cell dry weight, CDW) using a calibration curve as follows:

$$CDW_{CU} = \begin{cases} 22.187 \cdot CU - 5.0991 & CU \geq 0.9 \\ 11.124 \cdot CU + 0.66116 & CU < 0.9 \end{cases}$$

On-line CO_2 signals were obtained by measurements of the composition in the outflow (0%–25%). The volume measurement V was calculated by integration of the *on-line* flow signals every 10 s. Moreover, to consider the amount of volume taken for the *off-line* samples (8 mL/sample), these values were iteratively integrated with the pump signals. Although available every 10 s, each of the measurements was only used by the estimator every 60 s. This is done to not increase the size of the optimization problem.

4. Background theory

This section presents some important concepts utilized to develop the proposed approaches presented in the following section. We first introduce the sensitivity matrix (Section 4.1) which is the base for the regularization methods implemented. Than, briefly discuss the local observability and identifiability of the system (Section 4.2), for the further introduction of the stopping criterion used in the regularization methods.

4.1. Sensitivity matrix

The sensitivity matrix of model outputs is typically defined as the Jacobian of the outputs (y) with respect to the parameters θ (Vajda et al., 1985; Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009; Bae et al., 2021; Chen et al., 2022). However, because in this work we are interested in the effect of the parameters on the states, following the work of Bae et al. (2021), we define it as the Jacobian of the state variables (x) with respect to the parameters (θ):

$$S_{\theta,mn} = \left.\frac{\partial x_m}{\partial \theta_n}\right|_{\hat{x},\hat{\theta}} \tag{12}$$

where:

$m = 1, \dots, n_x$

$n = 1, \dots, n_\theta$

and $\hat{x}$ and $\hat{\theta}$ are respectively the estimates of states and parameters. The idea behind using the state variables (x) instead of the output measurements (y) is to identify the model parameters (θ) which have a significant effect on the state variables (Bae et al., 2021). This allows us to identify time-varying values for the parameters (θ) so that the estimated state variables ($\hat{x}$) can better match the *off-line* samples (y^*) and therefore the underlying true process behavior (i.e. obtain lower RMSE). The sensitivity matrix therefore summarizes locally the influence of the parameters on the state variables. It is important to carefully scale S_θ since this will directly affect the information obtained for the parameter estimation method (Vajda et al., 1989; Yao et al., 2003; Lund and Foss, 2008; Thompson et al., 2009; McLean and McAuley, 2012; Chis et al., 2016). Therefore, to address the problem of different orders of magnitudes in the parameters, their logarithm ($log(\theta)$) is used to avoid scaling issues due to differences in their orders of magnitude (Vajda et al., 1989; Chis et al., 2016). Additionally, to take into account the characteristics of the data and the available variation ranges, we will refer to the same scaling method applied in Thompson et al. (2009), McLean and McAuley (2012) and Bae et al. (2021), by using the state and parameter variance obtained from the covariance matrix. The scaled sensitivity matrix $Z \in \mathbb{R}^{n_x \times n_\theta}$ is defined as:

$$Z_{m,n} = \left.\frac{\partial x_m}{\partial log(\theta_n)} \frac{log(\sigma_{\theta_n})}{\sigma_{x_m}}\right|_{\hat{x},\hat{\theta}} \tag{13}$$

In the following, unless differently stated, we will refer to the sensitivity as defined in Eq. (13).

4.2. Local observability and identifiability

Observability of nonlinear systems based on *Lie* algebra is a structural and local property (Hermann and Krener, 1977; Isidori, 1985; Nijmeijer and van der Schaft, 1990; Powel and Morgansen, 2015; Villaverde, 2019). The local observability of the system in Eq. (1) was already analyzed and discussed by Tuveri et al. (2021). However, the focus there was limited to assess the possibility to infer information about the unmeasured variable of interest (S). Here instead, we take into account also the process parameters (θ) and therefore need to evaluate structural identifiability (Villaverde, 2019). By assuming that the only input is F_{in} (previously defined in Eq. (2)), being S_{in} and q_{air} constant over time (Section 2), the system in Eq. (5) can be written as an input-affine system, with the terms only dependent on the states ($\bar{f}_x$) and the ones dependent on the input ($\bar{f}_u$) such that:

$$\dot{x}_a = \bar{f}_x(x_a) + \bar{f}_u(x_a) \cdot u \tag{14}$$

$$y = h(x) \tag{15}$$

If we now consider that $\bar{f}_u(x_a)$ does not carry any information about the parameters (see Eq. (1)), we can define the map O, which represents the observation space of the system (Eq. (14)), without considering the term $\bar{f}_u(x_a) \cdot u$, as:

$$O = \{O_1, O_2, O_3\} = \{L^0_{\bar{f}_x} h, L^1_{\bar{f}_x} h, L^2_{\bar{f}_x} h\} \tag{16}$$

where:

$$L^0_{\bar{f}_x} h = h, \qquad L^1_{\bar{f}_x} h = \frac{\partial h}{\partial x_a} \bar{f}_x \qquad L^2_{\bar{f}_x} h = \frac{\partial (L^1_{\bar{f}_x} h)}{\partial x_a} \bar{f}_x$$

And obtain the codistribution dO:

$$dO = span\{d(O_1), d(O_2), d(O_3)\} \tag{17}$$

The dimension of the codistribution dO defines if the system is locally observable ($dO = n_x$). Moreover, when considering structural identifiability, if dO has the same size as the augmented vector ($n_x + n_\theta$), the system is said to be locally observable with identifiable parameters (Villaverde, 2019). The system in consideration (Eq. (1)) is locally observable, since $dO = n_x$ when considering only the states x, but not all the parameters are locally identifiable at every iteration, since $dO < n_x + n_\theta$ when considering the augmented vector x_a. Therefore indicating that the available information in the latter is not enough to estimate all states and parameters together.

5. Proposed approach

Simultaneous estimation of both states and parameters can lead to an ill-posed optimization problem. To attenuate the ill-posed problem

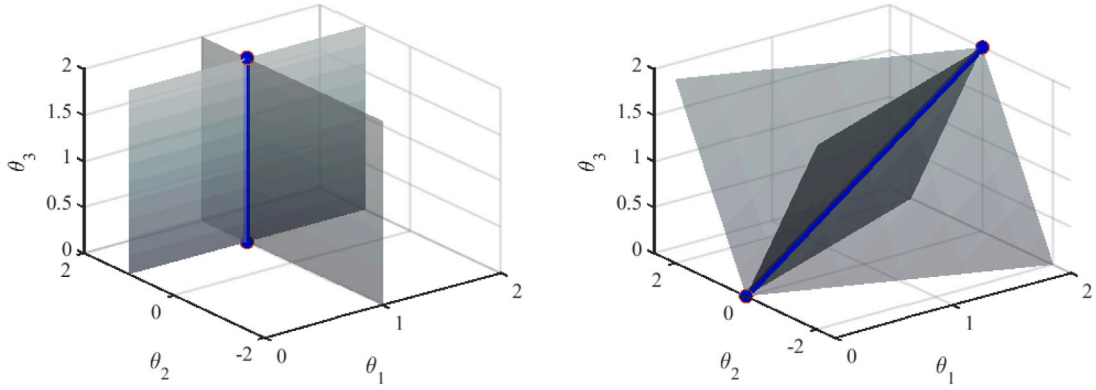

Fig. 1. Geometric representation of the constraint implementation. Orthogonalization Method (OM, left) and Subset Selection by Transformation (SST, right). Each plane represents an individual constrained region. When more than one constraint is selected, the constrained region is represented by their intersection (blue line). In OM (left), constraints are used for singular parameters (planes) and the degrees of freedom are the unconstrained parameters (blue line). In SST (right), constraints are implemented by linear combinations of different θ_i and therefore the planes are not required to be normal to an axis, leaving more freedom to the individual parameters. Here, both θ_1 and θ_3 (right) can vary within the projection of the constrained region (blue line) along their axes. The values reported here are only for illustration purposes.

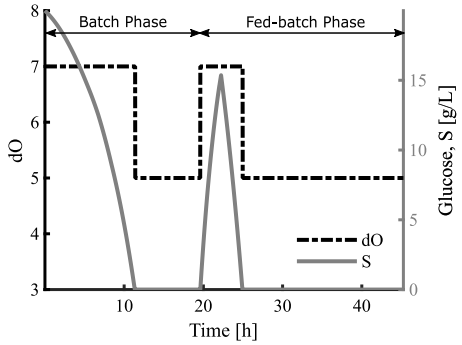

Fig. 2. Dimension of the observability space (dO) for the open loop model and Glucose (S) dynamics. dO is calculated without taking into account the terms related to the input (u, Eq. (14)). The decrease in dO ($dO = 5$) is a consequence of the sugar depletion.

that arises by the addition of the parameters as decision variables, we consider two different approaches, namely the Orthogonalization Method (OM) (Section 5.2.1) and the Subset Selection by Transformation (SST) (Section 5.2.2). Both methods are based on the sensitivity matrix and reduce the parameter space, but while OM keeps a subset of parameters constant, SST fixes linear combinations of parameters (i.e. clusters of parameters). The SST method can be therefore regarded as a more flexible approach, since it constrains linear combinations and not single parameters. This enables to vary more parameters simultaneously. However, a stopping criterion for the decision on how many degrees of freedom can be used for the estimation of the parameters is necessary for both methods (OM and SST). In this work we propose a stopping criterion based on the structural identifiability (Section 5.1).

This section will first present the stopping criterion proposed (Section 5.1), explaining how to practically use identifiability as a decision making criterion for constraints selection. Secondly, it introduces the two regularization approaches (i.e. OM and SST) together with the application of the proposed stopping criterion (Section 5.2). Thirdly, it shows the implementation of the aforementioned methods in a MHE formulation (Section 5.3), for the combined estimation of states and parameters.

5.1. Lie-based constraints selection

The intention of this work is to present a consistent and simple method to select the constraints for regularizing the optimization problem (Fig. 1). Although different regularization methods have been presented (McLean and McAuley, 2012), their dependence on a cut-off or threshold value might hinder their results (Chu et al., 2009; Wu et al., 2011; Kim and Lee, 2019; Nakama et al., 2020; Chen et al., 2022). We therefore want to use the information obtained by the local identifiability (structural property) as a baseline for the decision method of constraints selection. That is, at any time, we select the available degrees of freedom given the dimension of the observability co-distribution dO for the augmented system. The open loop test in Fig. 2 reports the dimension of the observability co-distribution dO for the system in Eq. (1), calculated as in Section 4.2. The system is locally observable ($dO > 4$, Fig. 2), but not all parameters are structurally identifiable ($dO < 7$ under sugar depletion, Fig. 2). However some information can still be retrieved also when the substrate is depleted. During the dynamic phases of the process (0–11 h and 20–25 h) the available information from the data is higher. Both under absence ($dO = 7$) and presence ($dO = 5$) of sugar depletion, it is possible to infer information of some of the parameters without having an ill-posed problem.

To introduce the method, we start by defining the dimension of the observability co-distribution dO. This is equivalent to the maximum number of degrees of freedom (i.e. decision variables of the optimization problem):

$$\mathrm{DOF}_{max} = dO \tag{18}$$

Eq. (18) indicates that dO, being a local property, gives us information about the maximum number of states and parameters that can be estimated, given the available information at each time. The maximum DOF (DOF_{max}) is given by $n_x + n_y$, which is 7 for this problem. Four of these degrees of freedom ($\mathrm{DOF}_x = 4$) are used for the estimation of the state variables (n_x). Following the previous considerations, the degrees of freedom (DOF_θ) available for estimating the parameters are given by:

$$\mathrm{DOF}_\theta = \mathrm{DOF}_{max} - \mathrm{DOF}_x \tag{19}$$

Therefore, the number of constraints (n_{θ_s}) that we need to impose on the estimation problem to achieve identifiability is:

$$n_{\theta_s} = n_\theta - \mathrm{DOF}_\theta \tag{20}$$

That is, whenever $\mathrm{DOF}_\theta < n_\theta$, we will add linear constraints on the parameters to attenuate the ill-posed estimation problem. However, since in real microbial cultivations the substrate is not entirely consumed ($S \neq 0$), some precautions must be taken. For numerical reasons, small values of S (in this work $< 10^{-3}$ g/L) should be considered $S = 0$ during the *on-line* calculation of dO.

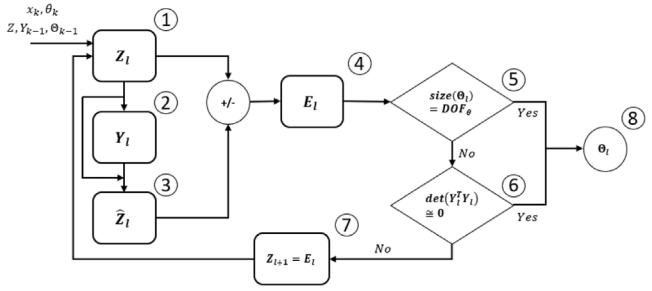

Fig. 3. Orthogonalization Method implementation block diagram. Initial values are given and the scaled sensitivity matrix is obtained in (1). Then the scaled parameter with the biggest norm is added to vector Y_l (2) and the matrix projection is calculated in (3), given (1) and (2). Consecutively, the residual error is obtained in (4). Finally, the stopping criteria are verified in (5) and (6). If one of these conditions is satisfied, the vector of constrained parameters is obtained in (8). More specifically, condition (6) is set with a threshold of 10^{-8}. If not the method goes to (7) and starts again to add an extra parameter to be constrained.

5.2. Regularization methods

This section describes the two regularization methods adopted (i.e. OM and SST) (Sections 5.2.1 and 5.2.2), together with their geometrical interpretation (Section 5.2.3).

5.2.1. Orthogonalization Method

Among the various Parameter Subset Selection (PSS) methods, the Orthogonalization Method is the heuristic method most commonly used due to its simplicity (Yao et al., 2003; Lund and Foss, 2008; Kim and Lee, 2019; Bae et al., 2021). In the context of MHE, this method is appealing due to its small computational loads (Bae et al., 2021). However, an unsuitable selection of the cut-off value for the stopping criterion, might result in poor estimates. To avoid that, we present here its implementation (Fig. 3) together with the stopping criterion previously presented (Section 5.1). For that purpose, we start with the scaled sensitivity matrix Z (Eq. (13)):

$$Z = [z_1,\ z_2, \dots,\ z_{n_\theta}] \tag{21}$$

The elements z_i of the Eq. (21) represent the column vectors of Z. Those vectors are sorted in decreasing order, by calculating the Euclidean norm. We take the one with the largest norm z_1, representing the main direction, and normalize it:

$$q_1 = \frac{z_1}{\|z_1\|} \tag{22}$$

At the first step ($l = 1$), $Z_l = Z$, $Y_{l-1} = \varnothing$ and $\Theta_{l-1} = \varnothing$. The parameter θ_l, corresponding to the column vector z_l, is than appended to the vector $\Theta_l = [\Theta_{l-1}, \theta_l]$. Similarly, the matrix Y_l is defined by appending q_l. Given Y_l, we calculate the matrix projection of Z_l onto it:

$$\hat{Z}_l = Y_l (Y_l^T Y_l)^{-1} Y_l^T Z_l \tag{23}$$

If we now subtract the projection $\hat{Z}_l$ (Eq. (23)) we obtain the residual error (Eq. (24)) which is orthogonal to Y_l:

$$E_l = Z_l - \hat{Z}_l \tag{24}$$

The method stops, differently from Bae et al. (2021), if the following condition, based on the stopping criterion defined in Section 5.1, is satisfied:

$$size(\Theta_l) = \mathrm{DOF}_\theta \tag{25}$$

where DOF_θ is obtained as in Eq. (19). If this condition is not satisfied, we set $Z_{l+1} = E_l$ until all the parameters are ranked or until $Y_l^T Y_l$ is nearly singular (Yao et al., 2003; Bae et al., 2021). The reader is referred to Strang (2016), Yao et al. (2003), Lund and Foss (2008), Thompson et al. (2009), McLean and McAuley (2012), Kim and Lee (2019) and Bae et al. (2021) for more detailed information.

5.2.2. Subset Selection by Transformation

Subset Selection by Transformation (SST, Fig. 4), was firstly introduced by Kim and Lee (2019) and further applied in Kim et al. (2019) and Chen et al. (2022). As observed in Nakama et al. (2020), and reported in Chen et al. (2022), Subset Selection by Transformation (SST) induces the same regularization effect as Principal Component Regression (PCR). Also known as Truncated Singular Values Decomposition (TSVD) (Kim and Lee, 2019), PCR is an eigenvalue–eigenvector based analysis that reveals existing dependencies among the parameters and that can reduce the parameter space by using the eigenvectors of the reduced Hessian matrix (Nakama et al., 2020). However, while in PCR the parameters are projected onto the subspace of the eigenvectors associated with the larger eigenvalues, in SST the parameter space is restricted by adding constraints in the directions of the eigenvectors associated with the smaller eigenvalues. Following the works of Kim and Lee (2019), Kim et al. (2019) and Chen et al. (2022), we linearized the model in Eq. (1) around $\theta = \theta_0$:

$$\begin{aligned} x &= x_0 + S_\theta(\theta - \theta_0) \\ &= S_\theta \theta + c \end{aligned} \tag{26}$$

where x_0 and θ_0 represent the initial values, S_θ is the sensitivity matrix (Eq. (12)) and $c = x_0 - S_\theta \theta_0$ is a constant. By adopting the Hessian approximation ($\mathcal{H} \approx S_\theta^T S_\theta$), as reported in Vajda et al. (1985), we can analyze the changes in the state variables x with respect to the parameters θ. Based on the eigenvalue–eigenvector decomposition, the Hessian approximation $\mathcal{H} \approx S_\theta^T S_\theta$, can be represented as:

$$S_\theta^T S_\theta = D \Sigma D^T \tag{27}$$

with $D \in \mathbb{R}^{n_\theta \times n_\theta}$ as the orthogonal eigenvectors matrix. Moreover, given that D is an orthonormal matrix ($D^T D = D D^T = \mathbb{I}$), we can rewrite Eq. (26) (Kim and Lee, 2019; Chen et al., 2022):

$$\begin{aligned} x &= S_\theta D D^T \theta + c \\ &= M\alpha + c \end{aligned} \tag{28}$$

where $M = S_\theta D$ and $\alpha = D^T \theta$. In this way we have obtained a vector $\alpha \in \mathbb{R}^{n_\theta}$ of transformed parameters. Note that the eigenvalue–eigenvector decomposition of the approximated Hessian based on either S_θ or Z (Eq. (13)), for the calculation of Eq. (28), does not change the direction of the eigenvectors (D), but just the magnitude of the eigenvalues (Ω). Eq. (27) can then be rewritten with respect to Eq. (13) as:

$$Z^T Z = D \Omega D^T \tag{29}$$

Consequently, based on the stopping criterion defined in Section 5.1, the number of active constraints (n_{θ_s}) (Eq. (20)) is defined. The vector α (Eq. (28)) can therefore be further divided in two subvectors α_1 and α_2 as:

$$x = M_1 \alpha_1 + M_2 \alpha_2 + c \tag{30}$$

where the magnitude of the eigenvalues sorts them in descending order. This two subvectors indicate the clusters of parameters corresponding to the larger eigenvalues ($\alpha_1 = D_1^T \theta$) which are estimated, and the ones corresponding to the smaller eigenvalues ($\alpha_2 = D_2^T \theta$) which are fixed (i.e. constraints). Given this decomposition, it is possible to find a constraint coefficient matrix $C \in \mathbb{R}^{n_{\theta S} \times n_\theta}$:

$$C\theta = r \tag{31}$$

where $C = D_2^T$ is a submatrix of D^T (Eq. (29)) and $r \in \mathbb{R}^{n_{\theta S}}$ is a given constraint right-hand side vector. By defining the constraint right-hand side vector as $r = C\theta_0$, Eq. (31) becomes:

$$C\theta - C\theta_0 = 0 \tag{32}$$

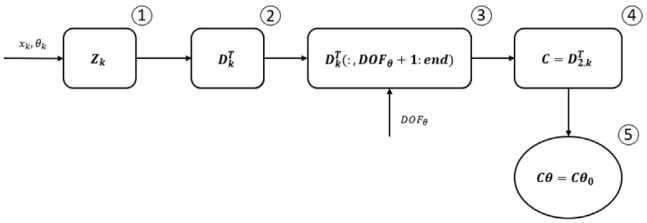

Fig. 4. Subset Selection by Transformation block diagram. This method requires the calculation of the right eigenvector matrix through singular values decomposition (SVD) in (2), given the scaled sensitivity matrix in (1). Then the submatrix of the right eigenvectors corresponding to the smallest eigenvalues is selected according to the DOF_θ at the k_{th} iteration (3) and set equal to the constraint coefficient matrix C (4). The constraints are selected in (5) and the right hand side defined based on the initial parameters θ_0.

Eq. (32) will therefore constrain, at each iteration k, the newly estimated parameters based on the nominal conditions (θ_0), which represent the local optimal solution (Table 1) previously obtained as discussed in Tuveri et al. (2021). Our contribution to the method, in addition to what previously presented in Kim and Lee (2019), consists in its geometric interpretation. Indeed, while the work of Kim and Lee (2019) looks at the problem through the transformation of the parameters (i.e. $\theta = D\alpha$, by rotating the axis through directions of the principal components) and constrains the unselected transformed parameters to their nominal values (i.e. $\alpha_2 = \alpha_{2,0}$), in our work we look at it from the constraints point of view (Nakama et al., 2020). Therefore showing that Eq. (32), which mathematically is equivalent to the method of Kim and Lee (2019), geometrically implies fixing the constraints in the directions of the smaller eigenvalues (Fig. 1) while maintaining the direction of the axes of the original parameters θ. This has the advantage of enhancing the interpretability of SST, which, as previously stated in Kim and Lee (2019), can complement OM and PCR. Indeed, our interpretation shows that: (1) by maintaining the same axes (Fig. 1), SST can be directly compared with the state of the art OM (Yao et al., 2003; Lund and Foss, 2008), as will also be further explained in Section 5.2.3; (2) by looking at the problem from the constraints point of view, SST has the same regularization effect as PCR.

5.2.3. Interpretation of regularization methods

Given the two regularization methods, OM (Section 5.2.1) and SST (Section 5.2.2), we want to show how the implementation of constraints differs. This concept is geometrically represented (for three dimensions) in Fig. 1, and mathematically translates, for OM, to:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{33}$$

and, for SST, to:

$$\begin{bmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \tag{34}$$

where Eqs. (33) and (34) are equivalent to Eq. (31). Indeed, while in OM we constrain single parameters θ at every iteration (Eq. (33)), in SST we constrain clusters of them (Eq. (34)). The main reasoning for using SST is to allow the change of the parameters along the directions with higher certainty, while restricting them along the less sensitive directions. These directions represent a linear combination (i.e. cluster) of parameters with a fixed relationship that can provide a more flexible approach than constraining single parameters (Fig. 1).

5.3. Moving Horizon Estimator

Consider the system dynamics, described by a set of ordinary differential equations (ODEs), which is given by the system in Eq. (1) augmented by considering the parameters as additional states (Eq. (5)) together with the measurement function $h(x) = Hx$ (Eq. (35b)). The implementation of one of the regularization methods previously presented OM (Section 5.2.1) and SST (Section 5.2.2) consists in adding a set of algebraic equations ($g(\theta) = 0$) to enforce equality constraints on the parameters (Eq. (35c)). We than obtain a system of differential algebraic equations (DAEs):

$$\dot{x}_a = \bar{f}(x_a, u) + w \tag{35a}$$

$$y = h(x) \tag{35b}$$

$$0 = g(\theta) \tag{35c}$$

where $x \in \mathbb{R}^{n_x}$ is the state vector, $y \in \mathbb{R}^{n_y}$ the output vector, $u \in \mathbb{R}^{n_u}$ the input vector, $\theta \in \mathbb{R}^{n_\theta}$ the parameters vector, $x_a \in \mathbb{R}^{(n_x+n_\theta)}$ the augmented vector and $w \in \mathbb{R}^{(n_x+n_\theta)}$ is the process noise random variable. Here w is different from $\omega \in \mathbb{R}^{(n_x+n_\theta)}$ which is solely a tuning parameter for $\bar{Q}_k$, indeed while $\omega \sim \mathcal{N}(0, Q_\omega)$, $w \sim \mathcal{N}(0, \bar{Q}_k)$. The discretization of Eq. (35), given the sampling time t_k, yields to:

$$x_{a,k+1} = \bar{F}(x_{a,k}, u_k) + w_k \tag{36a}$$

$$y_k = h(x_k) \tag{36b}$$

$$0 = g(\theta_k) \tag{36c}$$

$\bar{F} : \mathbb{R}^{(n_x+n_\theta)} \times \mathbb{R}^{n_u} \to \mathbb{R}^{(n_x+n_\theta)}$ is obtained by discretization of $\bar{f}$. The solution to the MHE problem (Kühl et al., 2011; Andersson et al., 2016) consists of finding parameters, states and their noise within a finite-time horizon $T = t_N - t_L$. Where t_L and t_N are respectively the initial and final times of the horizon. This is done by solving the following constrained least-squares optimization problem every 60 s (i.e. when new measurements are available), along the horizon T (30 min):

$$\min_{x_i, \theta_i, w_i} \left\| \begin{matrix} \hat{x}_L - x_L \\ \hat{\theta}_L - \theta_L \end{matrix} \right\|_{P_L}^2 + \sum_{i=L}^{N} \|y_i - h(x_i)\|_V^2 + \sum_{i=L}^{N-1} \|w_i\|_{W_k}^2$$

$$\text{s.t.} \quad x_{a,i+1} = F(x_{a,i}, u_i) + w_i \qquad i = L, \dots, N-1 \tag{37a}$$

$$g(\theta_i) = 0 \qquad i = L, \dots, N \tag{37b}$$

$$x_i \geq x_{min} \qquad i = L, \dots, N \tag{37c}$$

$$\theta_{min} \leq \theta_i \leq \theta_{max} \qquad i = L, \dots, N \tag{37d}$$

The aim is to obtain states (x_i), parameters (θ_i) and process noise (w_i) using the available information from the model (Eq. (1)) and the outputs (1 min sampling rate). The constraints in Eq. (37) are defined as:

- equality constraints on the state variables based on the process dynamics (Eq. (37a)) and on the process parameters, based on the regularization method selected in Eq. (37b) (Sections 5.2.1 and 5.2.2);
- inequality constraints based on physical limitations of the process variables ($x_{min} = [0, 0, 0, 0]$, Eq. (37c)) and the parameters uncertainty (given the parameter bounds θ_{min} and θ_{max}, Eq. (37d)). The latter allows to maintain the newly estimated parameters close to the nominal values, given that, under presence of structural plant-model mismatch, parameter adaptation might not necessarily lead to an improved model (Marchetti et al., 2008).

The cost function (Eq. (37)) is given by the summation of three Euclidean norms. The first term is the arrival cost which summarizes past information (its calculation is reported in S1). The second is the output noise cost, and the third is the process noise cost. The three terms are weighted by Kühl et al. (2011):

$$P_L = P^{-1/2}, \qquad V = R^{-1/2}, \qquad W_k = \bar{Q}_k^{-1/2} \tag{38}$$

where the notation for the squared norm is $\|b\|_B^2 = b^T B^T B b$ (Kühl et al., 2011). $P \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$, $R \in \mathbb{R}^{n_y \times n_y}$ and $\bar{Q}_k \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$

are respectively error, measurement noise and process noise (for the augmented system) covariance matrices. The terms $\hat{x}_L$ and $\hat{\theta}_L$ (Eq. (37)) represent, instead, the optimal estimates of x_L and θ_L.

The optimal solution of the nonlinear programming (NLP) was obtained using IPOPT (Wächter and Biegler, 2006) embedded in CasADi (Andersson et al., 2019), where the continuous time model was discretized by a three point Legendre collocation on finite elements.

6. Results

In the present section we first describe the implementation and the tuning parameters (Section 6.1). Then we present the results obtained. Firstly, to show how the presence of structural plant-model mismatch and the absence of full state feedback can hinder the estimation results, we present the possibility to use additional constraints on the parameters to find a trade-off between state and parameter estimates (Section 6.2). This is shown by using the Subset Selection by Transformation (SST) method, since one of the novelties of the work is to present SST as a more flexible regularization method in a MHE formulation. Consecutively, to show the advantages of SST, we compare the results obtained by its implementation to the ones obtained by the use of OM. To evaluate the accuracy of the estimates of biomass X and glucose S, and compare the results obtained through the different methods, we calculate the root mean squared error (RMSE) with respect to the reference *off-line* measurements, which were not used at any time by the estimator. In addition, to avoid that the zero values for the glucose could average out the RMSE, we calculate it separately for both batch (0–8 h) and fed-batch phase (20–30 h), without taking the zero values into account.

6.1. Implementation

The sampling rate was set to 1 min together with an estimation horizon T = 30 min. For the starting phase of the MHE problem, when the estimation window is smaller than the number of available measurements (i.e. $k < T$), the initial state vector was used (Kühl et al., 2011). The parameters θ, to be estimated along with the states, were appended to the augmented state vector x_a at each sampling interval. The MHE was initialized given the real values for the states and the nominal parameters (θ_0) reported in Table 1 as initial conditions:

$$x_{a,0} = [\,1.5\ \ 1.2\ \ 20\ \ 0\ \ 0.19445\ \ 0.007\ \ 0.006\ \ 0.42042\ \ 0.54308\,]^T$$

The initial error covariance matrix was selected as:

$$P_0^+ = \begin{bmatrix} P_{x,0}^+ & 0 \\ 0 & P_{\theta,0}^+ \end{bmatrix}$$

with $P_{x,0}^+$ obtained, according to Schneider and Georgakis (2013), as $P_0 = diag((\hat{x}_0 - x_0)^T(\hat{x}_0 - x_0))$:

$$P_{x,0}^+ = \begin{bmatrix} 2.09\cdot10^{-8} & 0 & 0 & 0 \\ 0 & 1.10\cdot10^{-5} & 0 & 0 \\ 0 & 0 & 1.09\cdot10^{-4} & 0 \\ 0 & 0 & 0 & 2.17\cdot10^{-5} \end{bmatrix}$$

and

$$P_{\theta,0}^+ = \begin{bmatrix} 1.05\cdot10^{-11} & 0 & 0 & 0 & 0 \\ 0 & 1.54\cdot10^{-11} & 0 & 0 & 0 \\ 0 & 0 & 2.02\cdot10^{-11} & 0 & 0 \\ 0 & 0 & 0 & 1.28\cdot10^{-11} & 0 \\ 0 & 0 & 0 & 0 & 4.95\cdot10^{-12} \end{bmatrix}$$

The measurement noise covariance matrix R:

$$R = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-1} & 0 \\ 0 & 0 & 10^{-3} \end{bmatrix}$$

And the process noise covariance matrix for the augmented system $\bar{Q}_k$ was tuned as in Section 3.1.1 (Eq. (11)), maintaining Q_ω and Q_θ unchanged throughout the different case studies. Moreover, for the arrival cost calculation (Sec. S1), to take into account the variation between the past information and the new horizon T, a weighting matrix $\bar{W}_L \in \mathbb{R}^{(n_x+n_\theta)\times(n_x+n_\theta)}$ (Sec. S1) is defined as follows:

$$\bar{W}_L = \begin{bmatrix} W_L & 0 \\ 0 & \mathbb{I} \end{bmatrix}$$

where $W_L \in \mathbb{R}^{n_x\times n_x}$ is defined as $W_k = Q_k^{-1/2}$ for $k = L$ and $\mathbb{I}^{n_\theta\times n_\theta}$ is the identity matrix. $\mathbb{I}$ was selected as the identity matrix so that, being always smaller than W_k, changes with respect to the past information in the states are penalized more than the ones in the parameters.

Table 3
Lower (θ_{min}) and upper (θ_{max}) parameter bounds.

Case	$\theta_{min}/\theta_{max}$	μ_{max}	K_s	k_d	Y_{XS}	$Y_X CO_2$
Case 1	θ_{min}	0.15	$5\cdot10^{-3}$	$1\cdot10^{-5}$	0.3846	0.10
	θ_{max}	0.30	$8\cdot10^{-3}$	$7\cdot10^{-3}$	0.4562	2
Case 2	θ_{min}	0.1941	$6.6\cdot10^{-3}$	$5.6\cdot10^{-3}$	0.4201	0.5429
	θ_{max}	0.1948	$7.4\cdot10^{-3}$	$6.4\cdot10^{-3}$	0.4208	0.5433
Case 3	θ_{min}	0.10	$5\cdot10^{-3}$	$1\cdot10^{-5}$	0.3846	0.10
	θ_{max}	0.40	$3\cdot10^{-2}$	$9\cdot10^{-3}$	0.4562	2

6.2. Trade-off between state and parameter estimates

To present the possibility to find a trade-off between model update and estimation accuracy, we present how the implementation of additional bounds on the parameters (i.e. inequality constraints, Eq. (37d)), differently leaves the possibility to adapt the model. In this extent, we present in this section the estimated states and the effects of the bounds on the parameters (i.e. inequality constraints) using three different cases. The aim is to show how the bounds selected can influence the performance of model prediction and estimation accuracy. Table 3 reports the bounds for the three presented cases.

6.2.1. Bounds selection

In *Case 1* the bounds on the parameters were selected to demonstrate the possibility to find a trade-off between model update and estimation accuracy (Table 3). This was achieved by carefully relaxing the standard deviations σ_{θ_i} reported in Table 1, therefore defining bounds on the parameters that leave them enough possibility to vary and adapt the model. Moreover, the bounds for Y_{XS} were kept narrower than for the other parameters because of its only dependence on the glucose (S).

Following the idea that an improvement in the model prediction can compromise the accuracy of the estimates, in *Case 2* we selected tighter bounds on the parameters (Table 3). The main idea is to allow the parameters to only vary within their uncertainty range as also reported in Kim et al. (2019), therefore reducing the penalization on the accuracy of the estimates. These bounds were also selected based on the standard deviations σ_{θ_i} reported in Table 1. However, considering that the standard deviation for the parameters was obtained by a single experiment, the values were relaxed by increasing the order of magnitude of the variance either by two times what reported in Table 1 (i.e. $\theta_0 \pm 10^2 \cdot \sigma_{\theta_i}$). This was done to take only batch-to-batch variations into account.

To show the influence and necessity of bounds to further regularize the optimization problem, in *Case 3* we selected higher variation margins for the parameters. Indeed, the bounds were defined to leave more freedom to the parameters to adapt, however, considering their necessity to avoid the deterioration of the glucose estimates. The only parameter maintained with tight bounds was Y_{XS}, due to its only dependence on S. Moreover, the reader is here warned that lower values were selected as non zero because of the presence of the logarithm in the calculation of the sensitivities (Eq. (13)).

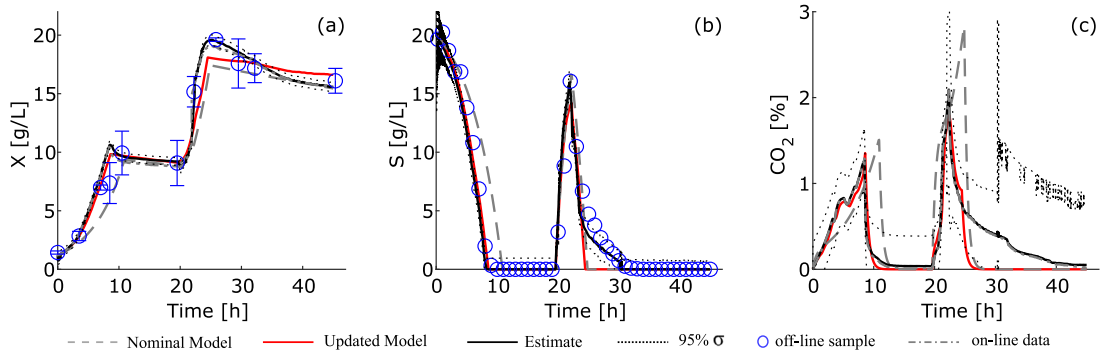

Fig. 5. Subset Selection by Transformation, *Case 1* - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). An improvement in the updated model is visible in (a) during the batch phase (0–20 h) and during the fed-batch phase (over 20 h). The updated model captures the sugar and the CO2 dynamics correctly until the end of the fed-batch phase (after 25 h). *Off-line* samples are only reported as reference, but not used by the estimator.

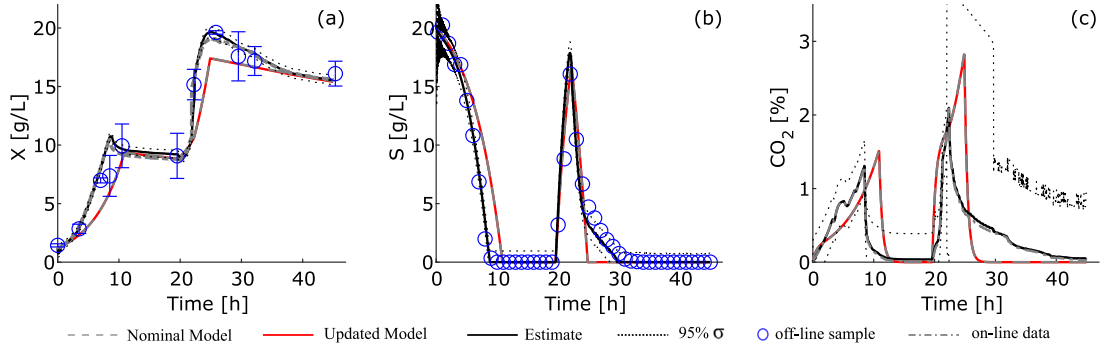

Fig. 6. Subset Selection by Transformation, *Case 2* - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). While improvement in the updated model is moderate, RMSE values show a good performance for the glucose estimate, with a value of 0.93 (b). *Off-line* samples are only reported as reference, but not used by the estimator.

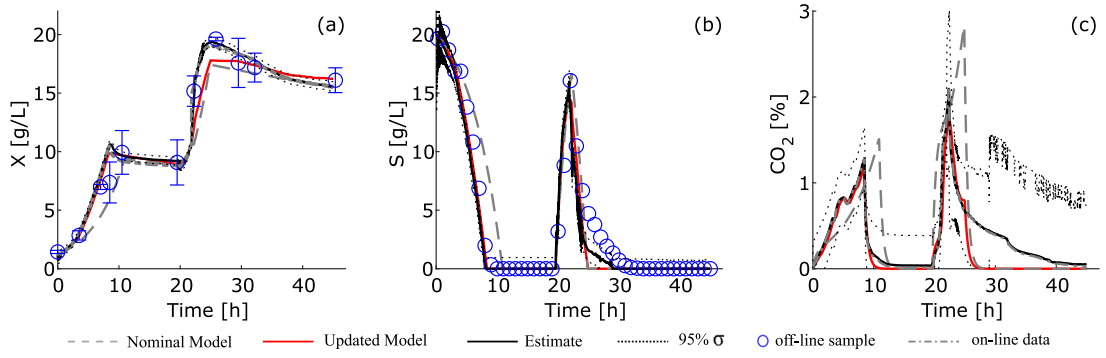

Fig. 7. Subset Selection by Transformation, *Case 3* - Biomass (a), glucose (b) and CO_2 (c) compositions and confidence intervals (95% σ). These results show the dependence of the methods on the bounds, to avoid deterioration of both state and parameter estimates. In (b) it is visible how the performance of the SST method is less influenced (in comparison to OM) by the choice of the bounds. *Off-line* samples are only reported as reference, but not used by the estimator.

6.2.2. *Estimation results using constrained regularization*

By solving the optimization problem, defined in Eq. (37), the estimates are:

- the *estimated states* ($\hat{x}$) presented in Figs. 5–7;
- the *estimated parameters* ($\hat{\theta}$, reported in S2), considered as additional states and estimated along with the states.

In addition, the results presented in Figs. 5–7 report:

- the *on-line data* (y) for X (y_X) and CO_2 (y_{CO_2}) which are the *on-line* measurements used by the estimator;

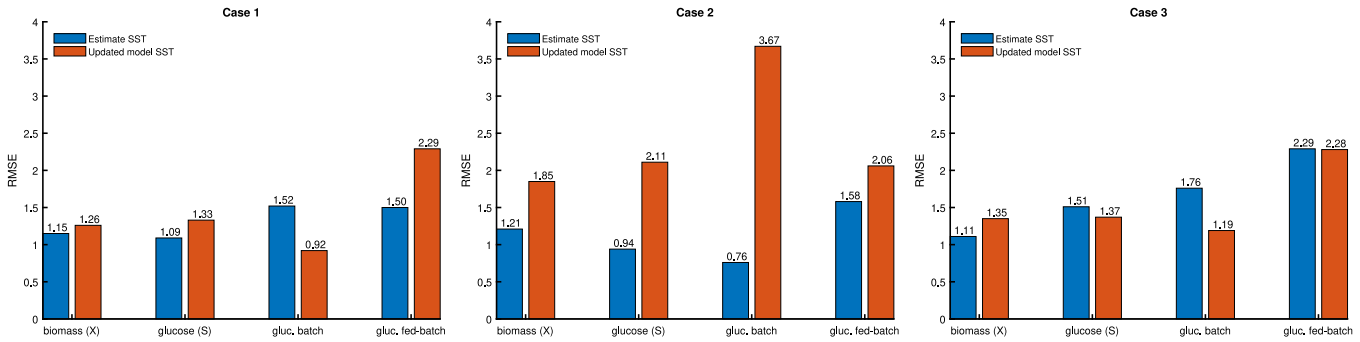

Fig. 8. Subset Selection by Transformation - RMSE values calculated for both the updated model and the estimated states with respect to the reference *off-line* values of biomass (X) and glucose (S). RMSE for glucose is also calculated for the single phases (batch and fed-batch) without taking the zero values into account.

Table 4
RMSE values calculated for both the nominal model and the estimated states with respect to the reference *off-line* values of biomass (X) and glucose (S) based on the results presented in Tuveri et al. (2022). RMSE for glucose is also calculated for the single phases (batch and fed-batch) without taking the zero values into account. The values of the parameters for the nominal model are reported in Table 1.

RMSE	Biomass	Glucose		
	tot.	tot.	Batch	Fed-batch
Estimate	1.23	1.01	0.78	1.71
Nom. Model	1.83	2.11	3.68	2.05

- *off-line samples* (y^*) for X ($y_X{}^*$) and S ($y_S{}^*$) which are only reported as reference values to validate the accuracy of the estimates and used to calculate the root mean squared error (RMSE), but not used at any time by the estimator;
- *confidence intervals* for the estimates (95% σ), obtained from the diagonal elements of the error covariance matrix P.

This results are obtained by the implementation of the MHE with SST as regularization method. To evaluate their accuracy, the RMSE (Fig. 8) is calculated with respect to the *off-line* measurements (y^*) of biomass X ($y_X{}^*$) and substrate S ($y_S{}^*$), reported by blue dots in Fig. 5a and b, as follows:

$$RMSE_{y^*} = \sqrt{\frac{1}{n_{y^*}} \sum_{j=1}^{n_{y^*}} (y^* - \hat{x})^2} \tag{39}$$

Additionally, to show the limitation of the model on describing the dynamic behavior of the state variables (x), due to the plant-model mismatch (particularly visible during the fed-batch phase), we report in Figs. 5–7 the model simulations for:

- the *nominal model* (i.e. open loop model), simulated by using the nominal parameters (θ_0), reported in Table 1;
- the *updated model*, simulated a posteriori, using the estimated parameters ($\hat{\theta}$), reported in S2.

Firstly, to give the reader a reference of the accuracy on the estimates obtained by simultaneously estimating states and parameters, we present the RMSE obtained by solely estimating the states, as previously presented in Tuveri et al. (2022) (Table 4).

Relatively accurate and comparable estimates were found for the biomass, with RMSE of 1.15 (*Case 1*, Fig. 5a), 1.21 (*Case 2*, Fig. 6a) and 1.11 (*Case 3*, Fig. 7a), presenting respectively 5%, 3% and 10% improvement with respect to solely estimating the states (Table 4). The estimates of glucose present instead a bigger difference, resulting in a RMSE of 1.09 for *Case 1*, 0.94 for *Case 2* and 1.53 for *Case 3*. This therefore results in the sole improvement of *Case 2*, which improves the estimate of the glucose by 7% with respect to solely estimating the states (Table 4), while *Case 1* and *Case 3* penalize it by 8% and 51% respectively. Moreover, the RMSE of the transient phases only (duration in which sugar values are not zero) shows that the error varies between 1.50 and 1.52 for *Case 1*, 0.76 to 1.58 for *Case 2* and 1.76 to 2.36 for *Case 3* (Fig. 8). Therefore, while *Case 2* shows an overall higher accuracy on the glucose estimates, *Case 1* presents a more balanced error between batch and fed-batch phase and *Case 3* presents the highest RMSE.

Effect on the unmeasured states. The results show that *Case 2*, by presenting tighter bounds, yields to more accurate estimate for the glucose (Fig. 6b). Due to the fact that both states and parameters are estimated in absence of direct glucose measurements, it is reasonable that the estimates will be a compromise between improved model prediction capabilities and accurate estimates. Indeed, parameters related to the glucose dynamics (i.e. μ_{max}, K_s and Y_{XS}, Eq. (1)) are dependant on the available information on the glucose. More specifically, Y_{XS}, which is only dependent on the glucose, is the most affected. Therefore, tighter bounds on these parameters lead to more accurate glucose estimates.

Trade-off between variance and bias. When comparing the results obtained in the three different cases, it is possible to see that the unmeasured state (i.e. glucose, S) presents an higher bias in *Case 1* and *Case 3* (Figs. 5b and 7b) with respect to *Case 2* (Fig. 6b). However, as mostly visible for the CO_2, *Case 2* presents higher variance in the estimates (Fig. 6c) with respect to the other two. Therefore, this interestingly shows how the combined estimation of state and parameters presents the necessity to find a trade-off between reducing variance and bias.

Effect of structural model mismatch. The updated model, calculated with the newly estimated parameters, yields to good predictions in *Case 1* ($RMSE_X = 1.26$ and $RMSE_S = 1.33$) and *Case 3* ($RMSE_X = 1.35$ and $RMSE_S = 1.38$), especially for the batch phase, presenting respectively an RMSE of 0.92 (Fig. 5b) and 1.18 (Fig. 7b) for the glucose. Differently, *Case 2*, having tighter parameters bounds, does not presents such improvement in the predictions as the other two cases ($RMSE_X = 1.84$ and $RMSE_S = 2.11$, Fig. 6). Instead it presents a slight penalization of the model prediction of less than 1% (Table 4). However, the improvement is limited in all the cases, due to the lack of knowledge on the process dynamics (i.e. structural plant-model mismatch, 25–30 h Figs. 5–7). It is sensible to understand that the improvement can not overpass the limitations of the model in use. In fact, the RMSE for the updated model, shows an increase of almost 40% between batch and fed-batch phases (Fig. 8), due to the higher structural plant-model mismatch in the latter. The updated parameters values are reported in Sec. S2.

6.3. Subset Selection by Transformation as an alternative to the Orthogonalization Method

To show the advantages of implementing SST in an MHE formulation, we compare the results previously presented (Section 6.2.2) to the ones obtained by the implementation of OM (Figs. 9–11).

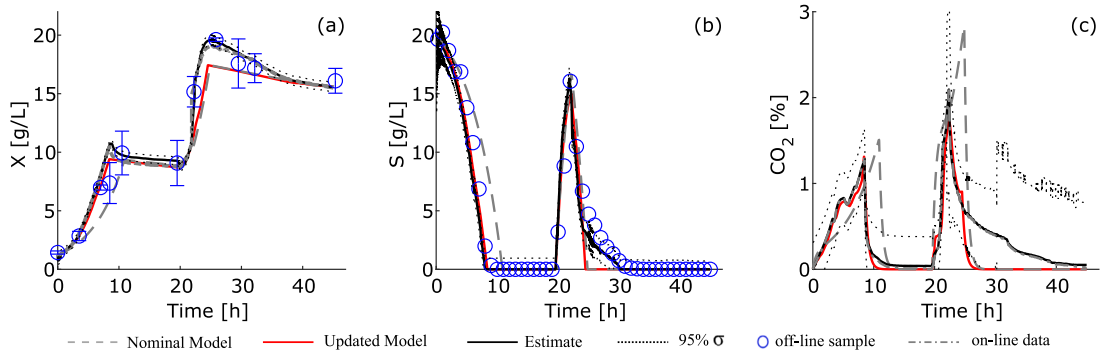

Fig. 9. Orthogonalization Method, Case 1 - Biomass (a), glucose (b) and CO_2 (c) compositions, and their confidence intervals (95% σ). An improvement in the updated model is visible in (a) in the batch (0–20 h) while less pronounced in the feeding phase (25–30 h). Similarly, the parameter adaptation could not yield to a glucose model (b) better than the nominal during the feeding phase. In contrast, the updated model for the CO_2 (c) could compensate better for the first 25 h. *Off-line* samples are only reported as reference, but not used by the estimator.

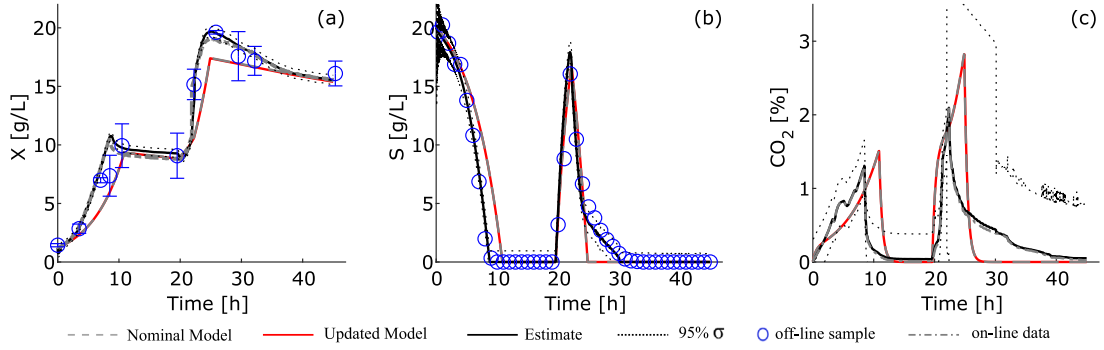

Fig. 10. Orthogonalization Method, Case 2 - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). While the improvement in the updated model is moderate, RMSE values show a good performance for the glucose estimate, with a value of 0.96 (b). *Off-line* samples are only reported as reference, but not used by the estimator.

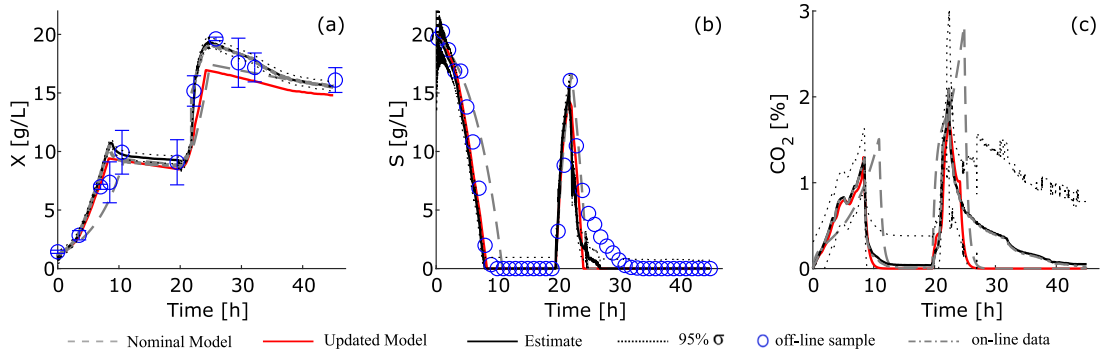

Fig. 11. Orthogonalization Method, Case 3 - Biomass (a), glucose (b) and CO_2 (c) compositions and their confidence intervals (95% σ). These results show the dependence of the methods on bounds to avoid deterioration of both states and parameter estimates. This is mostly noticeable in (b). *Off-line* samples are only reported as reference, but not used by the estimator.

Case 1. The estimates for biomass obtained by SST and OM are comparable, with RMSE of 1.15 for SST (Fig. 8) and 1.16 for OM (Fig. 12). The estimates of glucose present instead a difference, with a 18% increase in RMSE for OM (1.29) with respect to SST (RMSE 1.09). Moreover, the difference increases when looking only at the transient phases (1.66–1.96 for OM and 1.50–1.52 for SST). Additionally, also the updated model presents a 30% improvement in the predictions by SST (RMSE 0.92) with respect to OM (RMSE 1.26).

Case 2. The RMSE values (Figs. 8 and 12) show similar results, especially for the glucose estimates (RMSE 0.94 for SST and 0.91 for

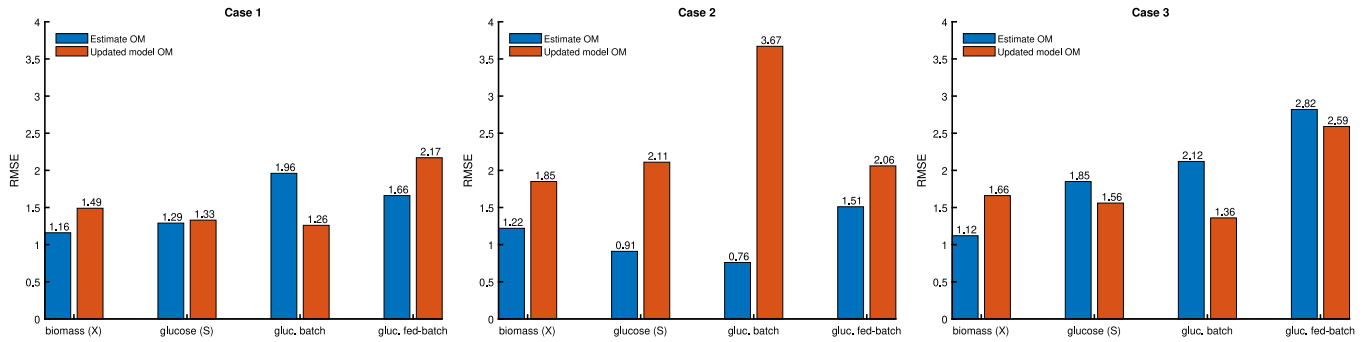

Fig. 12. Orthogonalization Method - RMSE values calculated for both the updated model and the estimated states with respect to the reference *off-line* values of biomass (X) and glucose (S). RMSE for glucose is also calculated for the single phases (batch and fed-batch) without taking the zero values into account.

OM). These results show that, when adopting tighter bounds on the parameters, the performance of the two methods is similar. However, it is important to note that while OM presents a slightly better improved RMSE for the glucose estimate (0.91) while presenting an RMSE for the biomass estimates of 1.22, the SST presents a more pronounced improvement in both of them (0.94 and 1.21) with respect to the case when only estimating the states (Table 4).

Case 3. The results obtained in this case (Fig. 11) show that OM is more influenced than SST (Fig. 7) by the selection of the bounds. Indeed, the RMSE calculated for the glucose estimates and the updated model in the batch phase are respectively 15 and 20% higher for OM (Fig. 12).

These results show us how both methods necessitate bounds to limit the possibility of deteriorating both state and parameter estimates. However, they also show that SST, by yielding lower RMSE values under looser bounds, presents less dependence on the bounds selection.

7. Discussion

In this work we present the use of two different methods, namely Orthogonalization Method (OM) and Subset Selection by Transformation (SST), for combined state and parameter estimation in a MHE framework, where the estimates were obtained based on the knowledge of the *on-line* measurements of volume (V), biomass (X) and carbon dioxide (CO_2).

One of the main contributions of this work is to apply MHE as a state estimator, to an experimental data-set, to estimate, in addition to the states, the model parameters. This is done to account for model uncertainty in the experimental data-set. However, as visible when comparing the RMSE values (Fig. 8), the more uncertain the model parameters are (i.e. model uncertainty), the larger is the estimation bias of the unmeasured states. This is consistent with what is stated in Dochain (2003). Moreover, when including all the parameters, an ill-posed problem arises. To alleviate that, we implemented and compared two regularization methods (i.e. OM and SST). Therefore taking practical identifiability into consideration (McLean and McAuley, 2012; Chis et al., 2016; Kim and Lee, 2019). The advantage of this implementation is to enable the selection of different subsets of parameters during the process, to avoid overfitting when adapting them. Another novelty of this work, therefore consists in the implementation of the SST method in an MHE and its comparison to the performance with a state of the art method as the OM. Additionally, an important aspect of this work is also the introduction of a stopping criterion based on the structural identifiability of the system (Villaverde, 2019). This bypasses the need to define an heuristic threshold value for the selection of the number of active constraints, therefore leading to a more consistent model adaptation. Moreover, the choice of this stopping criterion allows both methods to have a comparable way to select the available degrees of freedom. Additional stopping criteria are available in literature, like the definition of a threshold as in Nakama et al. (2020) and Bae et al. (2021), or the minimization of the Mean Squared Error (MSE) proposed by Chu et al. (2009), Wu et al. (2011) and applied in Kim and Lee (2019) and Chen et al. (2022). Indeed, the comparison between the two regularization methods, when the stopping criteria is a threshold, would be purely based on trial and error tuning of this threshold, and will not necessarily tell us about how effective the two regularizations are. We did not compare the proposed criteria with the minimization of the MSE because, due to the absence of full-state feedback (i.e. measurements of the substrate S is missing) and therefore the lack of information on all the state variables, an *on-line* selection of the degrees of freedom would not be possible.

Our work presents the possibility to estimate glucose by *on-line* measurements of other states (i.e. volume, biomass and CO_2), whereas the estimation of the internal states (i.e. parameters) was not the main focus, the parameters were estimated in order to adapt the model uncertainty. Indeed, the idea is that parameters should adapt in order to consider batch-to-batch variation, aiming to obtain more accurate state estimates (i.e. lower RMSE with respect to the off-line samples y^*). Consistent with Bae et al. (2021), simultaneous estimation of states and parameters improves the estimates of the variables of interest.

The results obtained for *Case 1* in Figs. 5 and 9 indicate that the SST method yields improved accuracy over the OM method, for both glucose estimate and model prediction capabilities. This improved accuracy (see Section 6.3) is a consequence of the different parameter estimation strategies. Indeed, in all the cases OM allows to vary a maximum of three parameters per iteration (see Sec. S2). In contrast, in SST all the parameters can potentially vary, within the clusters. Therefore, enabling higher flexibility for model adaptation. This is in agreement with what is reported in Kim and Lee (2019), Nakama et al. (2020) and Chen et al. (2022), stating that transformation of the constraints provides more flexibility to the regularization. Additionally, the results of *Case 2* indicate that tighter bounds on the model parameters yield better accuracy of the glucose estimates, both for SST (RMSE 0.94) and OM (RMSE 0.91), reducing the effect of the bias. Therefore presenting a good compromise for the glucose estimates (Figs. 6b and 10b). Lastly, *Case 3* supports these results, since by relaxing the bounds the estimates exhibited deterioration in both states and parameters. Therefore showing that both methods necessitate additional bounds (i.e. inequality constraints). This necessity emerges mainly for two reasons: (1) The discontinuities in the derivatives, due to sudden changes in the process dynamics (e.g. 8, 20 and 23 h, Fig. 7), which cause the parameters to hit the bounds, and (2) The absence of knowledge about the dynamics of the process (i.e. structural plant-model mismatch), which causes limitations on the model improvement. However, the results presented in *Case 3* show that SST (Fig. 8), by maintaining a fixed relationship between the parameters, presents less dependence on the bounds in comparison to OM (Fig. 12), where the bounds are the only limitation on the parameters. Since, as also reported in Kim et al. (2019), the choice of the bounds reflects the level of uncertainty in the parameters, their selection can be based on prior knowledge or historical data

(i.e. $\theta_0 \pm \alpha \cdot \sigma_\theta$) depending on the level of conservativeness desired (i.e. *Case 2*) or by certain physical considerations (i.e. *Case 1* and *Case 3*).

At this point, it is possible to discuss the differences between OM and SST. In fact, the different selection of constraints on the parameters will change the search region of the optimization problem (i.e. differently regularize the problem, Fig. 1). For instance, SST will use a different orthonormal base for the constraint selection, in comparison to the OM, based on the eigenvectors of the approximated Hessian, creating clusters that maintain the same relation between the parameters as in the nominal case (θ_0). Conversely, OM will yield solutions that maintain singular parameters at nominal values (θ_0). Therefore, while SST allows more parameters to vary (i.e. clusters) but with fixed relationships among them, OM allows fewer parameters to vary freely (Fig. 1), and is therefore a less flexible approach. In other words, while OM reduces the number of decision variables, SST reduces their search region.

The optimization problem finds local optimal estimates (of states x and parameters θ) given the available information (i.e. model, measurements and constraints). This strongly depends on the available measurements and the regularization method selected, and therefore on the given constraints that intrinsically carry information (Psichogios and Ungar, 1992). Indeed, due to the absence of multi-rate full state feedback, also the unmeasured states are affected, potentially presenting a large bias in the estimates (Dochain, 2003). Under this scenario, the available information over a finite horizon is limited. As a consequence, the model is likely to overfit the data, therefore resulting in poor model predictions (Bae et al., 2021). Additionally, as reported in Bonné and Jørgensen (2001), the use of regularization results in biased estimates, becoming a trade-off between reducing variance and bias. As shown in the results, *Case 1* and *Case 3* present an higher bias (Figs. 5b and 7b) but a lower variance (Figs. 5c and 7c) compared to *Case 2* (Fig. 6b and c). It is therefore important to consider that, given the total absence of information on the sugar measurements, limitations on the parameter estimation, and consequently on the prediction capability of the model will arise. In fact, as visible in the comparison of the three cases presented (Section 6.2.2), a high variation in the parameters related to the sugar consumption (i.e. μ_{max}, K_S and Y_{XS}), although could improve model prediction, would decrease the accuracy of the sugar estimates. Because these estimates are dependent on the tuning of the weights in the optimization (i.e. MHE), it is reasonable to assume that one possible solution is an iterative correction of the weight Q_k for the model error cost (Dochain, 2003), to account for improvement in the model prediction. This, which is also in agreement with the work of Kim et al. (2019), where the authors state that, to avoid biased estimates when unanticipated model mismatch occurs, significant a priori knowledge about the structural model mismatch is necessary, will be further investigated in future works.

Finally, it is important mentioning that the results obtained by SST are satisfactory, given the simplicity of the model and its limitations. For instance, either the estimates or the model predictions can be improved, without the guarantee of an accurate tracking of the drifting model parameters (Bae et al., 2021). It can therefore be considered that the improvements in the estimates are due to the reduction of the effect of the uncertainty caused by plant-model mismatch and unmodeled disturbances (Kim et al., 2019), rather than the accurate estimation of model parameters (Bae et al., 2021). Thus, the parameters can only improve the model accuracy but they can not compensate for the lack of knowledge about its dynamics. In fact, it is clear from Figs. 5a, 6a and 7a that the model is unable to properly describe the transition between growth phase and steady state for the biomass, especially after feeding. This applies also for the transient phase of the glucose (25-30 h, Figs. 5b, 6b and 7b). Indeed, despite the complexity of biological processes (Becker and Wittmann, 2012), bioprocesses are usually represented by oversimplified, unstructured Monod models. However, they are based on very simplistic representations of the cellular metabolism, by involving lumped parameters for the description of the intracellular phenomena (Jabarivelisdeh et al., 2020). To further illustrate that, we refer the reader to Section S3, which shows the difference between batch and fed-batch phases by the same model.

A possible solution to overcome this limitation is the use of hybrid (grey-box) models (Zendehboudi et al., 2018; Narayanan et al., 2019; Boiroux et al., 2019; Bradford et al., 2021). Some pioeneering works on the use of hybrid models to compensate for the lack of knowledge on the process dynamics can be found in process system engineering applications (Johansen and Foss, 1992), and more specifically for fed-batch bioreactors in Psichogios and Ungar (1992). As stated in Boiroux et al. (2019), hybrid models attempt to combine the advantages of first principle (i.e. white-box) and black-box (Narayanan et al., 2019) or Gaussian Process models (Bradford et al., 2021), by using the synergy between them (Narayanan et al., 2019). Indeed, in the case of absence of complete knowledge of the cell metabolism (i.e. lack of knowledge on the dynamics of the process), the hybrid model can correctly follow the physics of the process and therefore exhibit improved interpolation and extrapolation capabilities (Narayanan et al., 2019). These results therefore provide the opportunity to exploit the advantages of SST in the case of implementation of hybrid models.

8. Conclusions

This work presents the simultaneous estimation of states and parameters by an MHE in a microbial experimental fed-batch process. We here presented the comparison of two different regularization methods implemented in an MHE, for the selection of the additional decision variables (i.e. parameters or clusters). In addition, we proposed a stopping criterion based on structural identifiability to avoid the selection of cut-off values for the constraints selection. The results also present how, under presence of high model-mismatch, the necessity of additional inequality constraints (i.e. bounds on the parameters) is necessary when not full state feedback is available. Although under certain limitations, the results present accurate estimates and the possibility to adapt the model *on-line*.

CRediT authorship contribution statement

Andrea Tuveri: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Caroline S.M. Nakama:** Conceptualization, Resources, Writing – review & editing, Visualization, Supervision. **José Matias:** Conceptualization, Writing – review & editing, Visualization, Supervision. **Haakon Eng Holck:** Software, Writing – review & editing. **Johannes Jäschke:** Writing – review & editing, Visualization, Funding acquisition. **Lars Imsland:** Conceptualization, Resources, Writing – review & editing, Visualization, Supervision. **Nadav Bar:** Conceptualization, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

A. Tuveri is greatful to the colleagues Allyne M. dos Santos, Joakim R. Andersen, Rafael D. de Oliveira and Simen Bjorvand for useful discussion. The project was funded by the Bio Based Industries Joint Undertaking (JU) under the European Union's Horizon 2020 research and innovation programme under grant agreement N^o 790507. J. Matias acknowledges financial support from the Norwegian Research Council, SUBPRO, grant number: 237893. C. S. M. Nakama acknowledges financial support from the Norwegian Research Council through FRIPRO Project SensPATH.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.compchemeng.2023.108183.

References

Anane, E., López C, D.C., Barz, T., Sin, G., Gernaey, K.V., Neubauer, P., Cruz Bournazou, M.N., 2019. Output uncertainty of dynamic growth models: Effect of uncertain parameter estimates on model reliability. Biochem. Eng. J. 150, 107247. http://dx.doi.org/10.1016/j.bej.2019.107247.

Andersson, J.A.E., Gillis, J., Horn, G., Rawlings, J.B., Diehl, M., 2019. CasADi: a software framework for nonlinear optimization and optimal control. Math. Program. Comput. 11 (1), 1–36. http://dx.doi.org/10.1007/s12532-018-0139-4.

Andersson, L.E., Scibilia, F., Imsland, L., 2016. An estimation-forecast set-up for iceberg drift prediction. Cold Reg. Sci. & Technol. 131, 88–107. http://dx.doi.org/10.1016/j.coldregions.2016.08.001.

Bae, J., Kim, Y., Lee, J.M., 2021. Multirate moving horizon estimation combined with parameter subset selection. Comput. Chem. Eng. 147, 107253. http://dx.doi.org/10.1016/j.compchemeng.2021.107253.

Balsa-Canto, E., Rodriguez-Fernandez, M., Banga, J.R., 2007. Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation. J. Food Eng. 82 (2), 178–188. http://dx.doi.org/10.1016/j.jfoodeng.2007.02.006.

Becker, J., Wittmann, C., 2012. Bio-based production of chemicals, materials and fuels - Corynebacterium glutamicum as versatile cell factory. Curr. Opin. Biotechnol. 23 (4), 631–640. http://dx.doi.org/10.1016/j.copbio.2011.11.012.

Boiroux, D., Mahmoudi, Z., Jorgensen, J.B., 2019. Parameter estimation in type 1 diabetes models for model-based control applications. In: 2019 American Control Conference. ACC, IEEE, pp. 4112–4117. http://dx.doi.org/10.23919/ACC.2019.8814933.

Bonné, D., Jørgensen, S.B., 2001. Batch to batch improving control of yeast fermentation. In: Computer Aided Chemical Engineering, Vol. 9. Elsevier, pp. 621–626. http://dx.doi.org/10.1016/S1570-7946(01)80098-5.

Bonvin, D., Srinivasan, B., Ruppen, D., 2001. Dynamic Optimization in the Batch Chemical Industry. Technical Report.

Bradford, E., Imsland, L., Reble, M., del Rio-Chanona, E.A., 2021. Hybrid Gaussian process modeling applied to economic stochastic model predictive control of batch processes. In: Faulwasser, T., Müller, M.A., Worthmann, K. (Eds.), Recent Advances in Model Predictive Control. Springer International Publishing, Cham, pp. 191–218. http://dx.doi.org/10.1007/978-3-030-63281-6_8.

Chen, W., Wang, B., Biegler, L.T., 2022. Parameter estimation with improved model prediction for over-parametrized nonlinear systems. Comput. Chem. Eng. 157, 107601. http://dx.doi.org/10.1016/j.compchemeng.2021.107601.

Chis, O.-T., Villaverde, A.F., Banga, J.R., Balsa-Canto, E., 2016. On the relationship between sloppiness and identifiability. Math. Biosci. 282, 147–161. http://dx.doi.org/10.1016/j.mbs.2016.10.009.

Chu, Y., Huang, Z., Hahn, J., 2009. Improving prediction capabilities of complex dynamic models via parameter selection and estimation. Chem. Eng. Sci. 64 (19), 4178–4185. http://dx.doi.org/10.1016/j.ces.2009.06.057.

Dochain, D., 2003. State and parameter estimation in chemical and biochemical processes: a tutorial. J. Process Control 13 (8), 801–818. http://dx.doi.org/10.1016/S0959-1524(03)00026-X.

Doyle III, F.J., 1998. Nonlinear inferential control for process applications. J. Process Control 8 (5–6), 339–353. http://dx.doi.org/10.1016/S0959-1524(98)00015-8.

Elsheikh, M., Hille, R., Tatulea-Codrean, A., Krämer, S., 2021. A comparative review of multi-rate moving horizon estimation schemes for bioprocess applications. Comput. Chem. Eng. 146, 107219. http://dx.doi.org/10.1016/j.compchemeng.2020.107219.

Gatzke, E.P., Doyle III, F.J., 2002. Use of multiple models and qualitative knowledge for on-line moving horizon disturbance estimation and fault diagnosis. J. Process Control 12 (2), 339–352. http://dx.doi.org/10.1016/S0959-1524(01)00037-3.

Goffaux, G., Wouwer, A.V., 2008. Design of a robust nonlinear receding-horizon observer-Application to a biological system. IFAC Proc. Vol. 41 (2), 15553–15558.

Grossmann, I.E., Sargent, R.W.H., 1978. Optimum design of chemical plants with uncertain parameters. AIChE J. 24 (6), 1021–1028. http://dx.doi.org/10.1002/aic.690240612.

Hermann, R., Krener, A., 1977. Nonlinear controllability and observability. IEEE Trans. Automat. Control 22 (5), 728–740. http://dx.doi.org/10.1109/TAC.1977.1101601.

Hernández Rodríguez, T., Posch, C., Pörtner, R., Frahm, B., 2021. Dynamic parameter estimation and prediction over consecutive scales, based on moving horizon estimation: applied to an industrial cell culture seed train. Bioprocess Biosyst. Eng. 44 (4), 793–808. http://dx.doi.org/10.1007/s00449-020-02488-1.

Isidori, A., 1985. Nonlinear Control Systems: An Introduction. Springer.

Jabarivelisdeh, B., Carius, L., Findeisen, R., Waldherr, S., 2020. Adaptive predictive control of bioprocesses with constraint-based modeling and estimation. Comput. Chem. Eng. 135, 106744. http://dx.doi.org/10.1016/j.compchemeng.2020.106744.

Johansen, T.A., Foss, B.A., 1992. Representing and learning unmodeled dynamics with neural network memories. In: 1992 American Control Conference. IEEE, pp. 3037–3043. http://dx.doi.org/10.23919/ACC.1992.4792705.

Kim, B., Huusom, J.K., Lee, J.H., 2019. Robust batch-to-batch optimization with scenario adaptation. Ind. Eng. Chem. Res. 58 (30), 13664–13674. http://dx.doi.org/10.1021/acs.iecr.8b06233.

Kim, B., Lee, J.H., 2019. Parameter subset selection and biased estimation for a class of ill-conditioned estimation problems. J. Process Control 81, 65–75. http://dx.doi.org/10.1016/j.jprocont.2019.05.015.

Kolås, S., Foss, B.A., Schei, T.S., 2009. Noise modeling concepts in nonlinear state estimation. J. Process Control 19 (7), 1111–1125. http://dx.doi.org/10.1016/j.jprocont.2009.03.002.

Kravaris, C., Hahn, J., Chu, Y., 2013. Advances and selected recent developments in state and parameter estimation. Comput. Chem. Eng. 51, 111–123. http://dx.doi.org/10.1016/j.compchemeng.2012.06.001.

Kühl, P., Diehl, M., Kraus, T., Schlöder, J.P., Bock, H.G., 2011. A real-time algorithm for moving horizon state and parameter estimation. Comput. Chem. Eng. 35 (1), 71–83. http://dx.doi.org/10.1016/j.compchemeng.2010.07.012.

Küpper, A., Diehl, M., Schlöder, J.P., Bock, H.G., Engell, S., 2009. Efficient moving horizon state and parameter estimation for SMB processes. J. Process Control 19 (5), 785–802. http://dx.doi.org/10.1016/j.jprocont.2008.10.004.

Küpper, A., Wirsching, L., Diehl, M., Schlöder, J.P., Bock, H.G., Engell, S., 2010. Online identification of adsorption isotherms in SMB processes via efficient moving horizon state and parameter estimation. Comput. Chem. Eng. 34 (12), 1969–1983. http://dx.doi.org/10.1016/j.compchemeng.2010.07.005.

Leu, G., Baratti, R., 2000. An extended Kalman filtering approach with a criterion to set its tuning parameters; application to a catalytic reactor. Comput. Chem. Eng. 23 (11–12), 1839–1849. http://dx.doi.org/10.1016/S0098-1354(00)00298-2.

Liu, J., Gnanasekar, A., Zhang, Y., Bo, S., Liu, J., Hu, J., Zou, T., 2021. Simultaneous state and parameter estimation: The role of sensitivity analysis. Ind. Eng. Chem. Res. 60 (7), 2971–2982. http://dx.doi.org/10.1021/acs.iecr.0c03793.

López C, D.C., Barz, T., Körkel, S., Wozny, G., 2015. Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. Comput. Chem. Eng. 77, 24–42. http://dx.doi.org/10.1016/j.compchemeng.2015.03.002.

Lund, B.F., Foss, B.A., 2008. Parameter ranking by orthogonalization—Applied to nonlinear mechanistic models. Automatica 44 (1), 278–281. http://dx.doi.org/10.1016/j.automatica.2007.04.006.

Marchetti, A., Chachuat, B., Bonvin, D., 2008. Real-time optimization via adaptation and control of the constraints. In: Computer Aided Chemical Engineering, Vol. 25. Elsevier, pp. 393–398. http://dx.doi.org/10.1016/S1570-7946(08)80070-3.

McLean, K.A.P., McAuley, K.B., 2012. Mathematical modelling of chemical processes-obtaining the best model predictions and parameter estimates using identifiability and estimability procedures. Can. J. Chem. Eng. 90 (2), 351–366. http://dx.doi.org/10.1002/cjce.20660.

Mohd Ali, J., Ha Hoang, N., Hussain, M., Dochain, D., 2015. Review and classification of recent observers applied in chemical process systems. Comput. Chem. Eng. 76, 27–41. http://dx.doi.org/10.1016/j.compchemeng.2015.01.019.

Monod, J., 1949. The growth of bacterial cultures. Annu. Rev. Microbiol. 3 (1), 371–394. http://dx.doi.org/10.1146/annurev.mi.03.100149.002103.

Nakama, C.S., Le Roux, G.A., Zavala, V.M., 2020. Optimal constraint-based regularization for parameter estimation problems. Comput. Chem. Eng. 139, 106873. http://dx.doi.org/10.1016/j.compchemeng.2020.106873.

Narayanan, H., Sokolov, M., Morbidelli, M., Butté, A., 2019. A new generation of predictive models: The added value of hybrid models for manufacturing processes of therapeutic proteins. Biotechnol. Bioeng. 116 (10), 2540–2549. http://dx.doi.org/10.1002/bit.27097.

Nijmeijer, H., van der Schaft, A., 1990. Nonlinear Dynamical Control Systems, Vol. 175. Springer New York, New York, NY, http://dx.doi.org/10.1007/978-1-4757-2101-0.

Powel, N.D., Morgansen, K.A., 2015. Empirical observability Gramian rank condition for weak observability of nonlinear systems with control. In: 2015 54th IEEE Conference on Decision and Control. CDC, IEEE, pp. 6342–6348. http://dx.doi.org/10.1109/CDC.2015.7403218.

Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. AIChE J. 38 (10), 1499–1511. http://dx.doi.org/10.1002/aic.690381003.

Raïssi, T., Ramdani, N., Candau, Y., 2005. Bounded error moving horizon state estimator for non-linear continuous-time systems: application to a bioprocess system. J. Process Control 15 (5), 537–545. http://dx.doi.org/10.1016/j.jprocont.2004.10.002.

Rao, C.V., 2000. Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete-Time Systems. The University of Wisconsin-Madison.

Rao, C.V., Rawlings, J.B., Mayne, D.Q., 2003. Constrained state estimation for nonlinear discrete-time systems: stability and moving horizon approximations. IEEE Trans. Automat. Control 48 (2), 246–258. http://dx.doi.org/10.1109/TAC.2002.808470.

Robertson, D.G., Lee, J.H., Rawlings, J.B., 1996. A moving horizon-based approach for least-squares estimation. AIChE J. 42 (8), 2209–2224. http://dx.doi.org/10.1002/aic.690420811.

Schei, T.S., 2008. On-line estimation for process control and optimization applications. J. Process Control 18 (9), 821–828. http://dx.doi.org/10.1016/j.jprocont.2008.06.014.

Schneider, R., Georgakis, C., 2013. How to NOT make the extended kalman filter fail. Ind. Eng. Chem. Res. 52 (9), 3354–3362. http://dx.doi.org/10.1021/ie300415d.

Strang, G., 2016. Introduction to Linear Algebra, fifth ed. Wellesley-Cambridge Press, pp. 206–240, Revised.

Thompson, D.E., McAuley, K.B., McLellan, P.J., 2009. Parameter estimation in a simplified MWD model for HDPE produced by a Ziegler-Natta catalyst. Macromol. React. Eng. 3 (4), 160–177. http://dx.doi.org/10.1002/mren.200800052.

Tuveri, A., Holck, H.E., Nakama, C.S., Matias, J., Jäschke, J., Imsland, L., Bar, N., 2022. Bioprocess monitoring: A moving horizon estimation experimental application. IFAC-PapersOnLine 55 (7), 222–227. http://dx.doi.org/10.1016/j.ifacol.2022.07.448.

Tuveri, A., Pérez-García, F., Lira-Parada, P.A., Imsland, L., Bar, N., 2021. Sensor fusion based on Extended and Unscented Kalman Filter for bioprocess monitoring. J. Process Control 106, 195–207. http://dx.doi.org/10.1016/j.jprocont.2021.09.005.

Vajda, S., Rabitz, H., Walter, E., LeCourtier, Y., 1989. Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chem. Eng. Commun. 83 (1), 191–219. http://dx.doi.org/10.1080/00986448908940662.

Vajda, S., Valko, P., Turányi, T., 1985. Principal component analysis of kinetic models. Int. J. Chem. Kinet. 17 (1), 55–81. http://dx.doi.org/10.1002/kin.550170107.

Valipour, M., Ricardez-Sandoval, L.A., 2021. Assessing the impact of EKF as the arrival cost in the moving horizon estimation under nonlinear model predictive control. Ind. Eng. Chem. Res. 60 (7), 2994–3012. http://dx.doi.org/10.1021/acs.iecr.0c06095.

Valipour, M., Ricardez-Sandoval, L.A., 2022a. A robust moving horizon estimation under unknown distributions of process or measurement noises. Comput. Chem. Eng. 157, 107620. http://dx.doi.org/10.1016/j.compchemeng.2021.107620.

Valipour, M., Ricardez-Sandoval, L.A., 2022b. Extended moving horizon estimation for chemical processes under non-Gaussian noises. AIChE J. 68 (3), e17545. http://dx.doi.org/10.1002/aic.17545.

Vercammen, D., Logist, F., Impe, J.V., 2016. Online moving horizon estimation of fluxes in metabolic reaction networks. J. Process Control 37, 1–20. http://dx.doi.org/10.1016/j.jprocont.2015.08.014.

Villaverde, A.F., 2019. Observability and structural identifiability of nonlinear biological systems. Complexity 2019, 1–12. http://dx.doi.org/10.1155/2019/8497093.

Wächter, A., Biegler, L.T., 2006. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. 106 (1), 25–57. http://dx.doi.org/10.1007/s10107-004-0559-y.

Wu, S., McLean, K.A., Harris, T.J., McAuley, K.B., 2011. Selection of optimal parameter set using estimability analysis and MSE-based model-selection criterion. Int. J. Adv. Mechatron. Syst. 3 (3), 188. http://dx.doi.org/10.1504/IJAMECHS.2011.042615.

Yao, K.Z., Shaw, B.M., Kou, B., McAuley, K.B., Bacon, D.W., 2003. Modeling ethylene/butene copolymerization with multi-site catalysts: Parameter estimability and experimental design. Polym. React. Eng. 11 (3), 563–588. http://dx.doi.org/10.1081/PRE-120024426.

Zendehboudi, S., Rezaei, N., Lohi, A., 2018. Applications of hybrid models in chemical, petroleum, and energy systems: A systematic review. Appl. Energy 228, 2539–2566. http://dx.doi.org/10.1016/j.apenergy.2018.06.051.