Proceedings of the 10th IFAC Conference on Foundations of Systems Biology in Engineering, Corfu Island, Greece, September 8-11, 2024

Robust Model Predictive Control of a Vaccine Production Unit

Micaela Benavides [*,**] **Laurent Dewasme** [*] **Pascal Gerkens** [**]
Gaël de Lannoy [**] **Alain Vande Wouwer** [*]

[*] *Systems, Estimation, Control and Optimization (SECO), University of Mons, 7000 Mons, Belgium, (e-mail:alain.vandewouwer@umons.ac.be)*
[**] *GSK, 1330 Rixensart, Belgium*

Abstract: In this paper, nonlinear model predictive controllers (NMPC) are proposed to optimize the biomass productivity of yeast fed-batch cultures. Their predictions are driven by a mechanistic model developed using a few industrial vaccine production data sets. The limited amount of data causes high parametric uncertainty levels and, to address this issue, a robust tube-based MPC is proposed and its robustness is assessed by a Monte-Carlo analysis, and compared to the classical MPC formulation.

Keywords: Mathematical model, parameter estimation, *Saccharomyces cerevisiae*, model predictive control, Biotechnology.

1. INTRODUCTION

Industrial vaccine manufacturing relies on recombinant protein production in bioreactors (Silva et al., 2022). Usually, this step is led by growing a genetically modified yeast host strain in fed-batch mode, i.e., by continuously feeding the cells until complete reactor filling and without medium withdrawal. This operation allows densification of the biomass concentration while limiting the quantity of consumables such as the feed medium.

This biomass densification is made delicate following the switching metabolism of yeast, function of the quantity of substrate in the reactor. Yeasts are indeed likely to follow two main metabolic pathways, exhibiting overflow metabolism (or short-term Crabtree effect, Crabtree (1929)) when the substrate is fed in excess or starving metabolism in the opposite case. Overflow metabolism leads to the accumulation of ethanol by fermentation, which tends to inhibit the yeast's oxygen capacity and, in turn, the biomass growth. The productivity optimization therefore requires tracking a singular and exponential feeding trajectory corresponding to the metabolic regime boundary (Sonnleitner and Käppeli, 1986).

An optimal feeding strategy, consisting of tracking the corresponding critical glucose concentration, can be established to avoid the undesired ethanol accumulation. Several studies have however shown that the potentialities of classical controllers were limited because the tracking of optimal growth conditions corresponds to an exponential trajectory (Akesson et al., 2001; Dewasme et al., 2011a). Moreover, biomass productivity is not the sole fundamental quality attribute, and its maximization is often combined with several other quality criteria constraining the optimization problem.

Economic Model predictive control (EMPC) therefore appears as a well-adapted solution dealing with process nonlinearities, possible singular trajectory tracking, as well as input and state variable constraints, while optimizing an economic objective function such as productivity, use of consumables or simply energy (Qin and Badgwell, 2000, 2003; Angeli et al., 2012).

Following the biological nature of the process, model parametric uncertainties also need to be considered and a robust MPC formulation is often necessary, either considering conservative approaches like min-max MPC (Santos et al., 2012; Dewasme et al., 2015), multi-stage MPC (Hebing et al., 2020; Dewasme et al., 2023), or tube-based MPC (Dewasme et al., 2024). These recent works highlight the effectiveness of the latter method in dealing with uncertain metabolic switches, corresponding to model structural changes that cannot be handled by adaptive systems (Dewasme et al., 2011b). The tube paradigm also presents an interesting analogy with the definition of quality-by-design bounds imposed in industrial pharmaceutical applications. Indeed, the property of the state and input trajectories to remain in a specific tube provides a useful prediction of these bounds.

This work aims to provide a robust tube EMPC formulation for yeast fed-batch cultures where a sole inlet feed rate is used to maximize biomass productivity. This paper is organized as follows. Section 2 presents the mechanistic model and the identified parameters using an industrial data set. Section 3 is dedicated to the classical MPC description, including the cost function and the control parameter design, while section 4 discusses the robustness of the classical MPC versus a robust tube-based formulation. Conclusions are drawn in section 5.

2. MODELING OF YEAST CULTURES

The dynamic model is inspired by the bottleneck assumption defined by Sonnleitner and Käppeli (1986), demonstrating that the yeast metabolism is ruled by its limited respiratory capacity, leading to substrate overflow (Crabtree, 1929). The proposed reaction scheme therefore counts three macroreactions describing substrate oxidation, substrate overflow, and ethanol oxidation, as follows:

$$\text{Substrate oxidation} : S \xrightarrow{r_1 X} k_1 X \tag{1a}$$

$$\text{Substrate overflow} : S \xrightarrow{r_2 X} k_2 E + k_4 X \tag{1b}$$

$$\text{Ethanol oxidation} : E \xrightarrow{r_3 X} k_3 X \tag{1c}$$

where S, X, and E respectively stand for substrate, biomass and ethanol concentrations.

Following the bottleneck assumption of Sonnleitner and Käppeli (1986), the yeast metabolism is likely to follow two main pathways depending on the available amount of substrate (glucose) and the corresponding respiratory capacity. When the latter is not filled, the yeasts are in respirative regime and the remaining respiratory capacity can be used to oxidize ethanol, therefore activating reactions (1a) and (1c) while (1b) is assumed to be negligible. However, if the respiratory capacity is filled, the excess of substrate which is not oxidized enters the fermentation pathway, producing ethanol. The corresponding fermentation regime considers the activations of only reactions (1a) and (1b) while (1c) is negligible.

Each reaction presents a rate of the following form:

$$r_1 = \mu_{m1} \cdot \frac{S}{K_{S1} + S} \cdot \frac{1}{1 + \frac{X}{K_{IX}}} \cdot \frac{1}{1 + \frac{E}{K_{IE}}} \tag{2a}$$

$$r_2 = \mu_{m2} \cdot \frac{S}{K_{S2} + S} \cdot \frac{1}{1 + \frac{X}{K_{IX}}} \tag{2b}$$

$$r_3 = \mu_{m3} \cdot \frac{E}{K_E + E} \cdot \frac{1}{1 + \frac{X}{K_{IX}}} \cdot \frac{1}{1 + \frac{S}{K_{IS}}} \tag{2c}$$

The first reaction rate r_1 (2a) considers Monod kinetics to describe glucose uptake, limited by the available respiratory capacity either inhibited by the biomass density (second factor) or the presence of ethanol (third factor). The second reaction rate r_2 (2b) is also ruled by a Monod factor related to glucose uptake and limited by the respiratory capacity depending on the biomass density. The third reaction rate r_3 (2c) is driven by the presence of ethanol using Monod kinetics, a respiratory capacity limitation factor comparable to reactions 1 and 2, and an inhibition factor by substrate, explaining the preferential selection, as main substrate, of glucose over ethanol. Applying mass balance to each macro-reaction yields the following differential equation system:

$$\frac{dX}{dt} = (k_1 \cdot r_1 + k_3 \cdot r_3 + k_4 \cdot r_2) \cdot X - D \cdot X \tag{3a}$$

$$\frac{dS}{dt} = -(r_1 + r_2) \cdot X - D \cdot S + D \cdot S_{in} \tag{3b}$$

$$\frac{dE}{dt} = (k_2 \cdot r_2 - r_3) \cdot X - D \cdot E \tag{3c}$$

$$\frac{dV}{dt} = D \cdot V = F_{in} \tag{3d}$$

where S_{in} represents the glucose concentration in the inlet feed, $D = F_{in}/V$ is the dilution rate, F_{in} is the inlet feed flow rate and V the bioreactor volume.

2.1 Experimental setup

In this study, a recombinant yeast strain of *Saccharomyces cerevisiae* is cultivated in fed-batch in a bioreactor with an initial volume of 5.5 L, and a stirrer speed initially set to 260 rpm. The temperature was maintained at 30°C throughout all the experimental sessions, while the pH was regulated up to 5, using a base solution. The bioreactor was equipped with an in-line pO2 sensor delivering dissolved oxygen measurement, a stirrer motor controlled to maintain this pO2 above 60%, and a peristaltic pump controlling the feed flow rate. The culture duration was set to 95 hours, and offline measurements of optical density (OD), glucose, ethanol, and ammonium were taken every 1 hour. The OD provides the concentration of biomass based on a dry weight calibration. Two different input (F_{in}) trajectories are considered. The first one is composed of pulses and the second one is exponential as shown in Fig. 1. The cultures were conducted in GSK laboratories, and any other detail remains confidential.

The parameter values estimated along with their respective coefficient of variation (CV), for the model previously described, are reported in Table 1.

Table 1. Parameter estimate values with their respective coefficient of variation (CV)

Parameter Name	Parameter Value	Units	CV(%)
μ_{m1}	1,343	[gS/gX/h]	8,0
μ_{m2}	0,208	[gS/g(X+E)/h]	51,4
μ_{m3}	0,700	[gE/gX/h]	23,5
K_{IX}	22,905	[gX/L]	2,2
K_{S1}	0,026	[gS/L]	28,9
K_{S2}	0,004	[gS/L]	41,4
K_{IE}	50,462	[gE/L]	31,8
K_{IS}	0,003	[gS/L]	42,0
K_E	0,158	[gE/L]	18,4
k_1	0,100	[g/g]	39,6
k_2	1,457	[g/g]	51,8
k_3	1,000	[g/g]	7,9
k_4	0,356	[g/g]	58,4

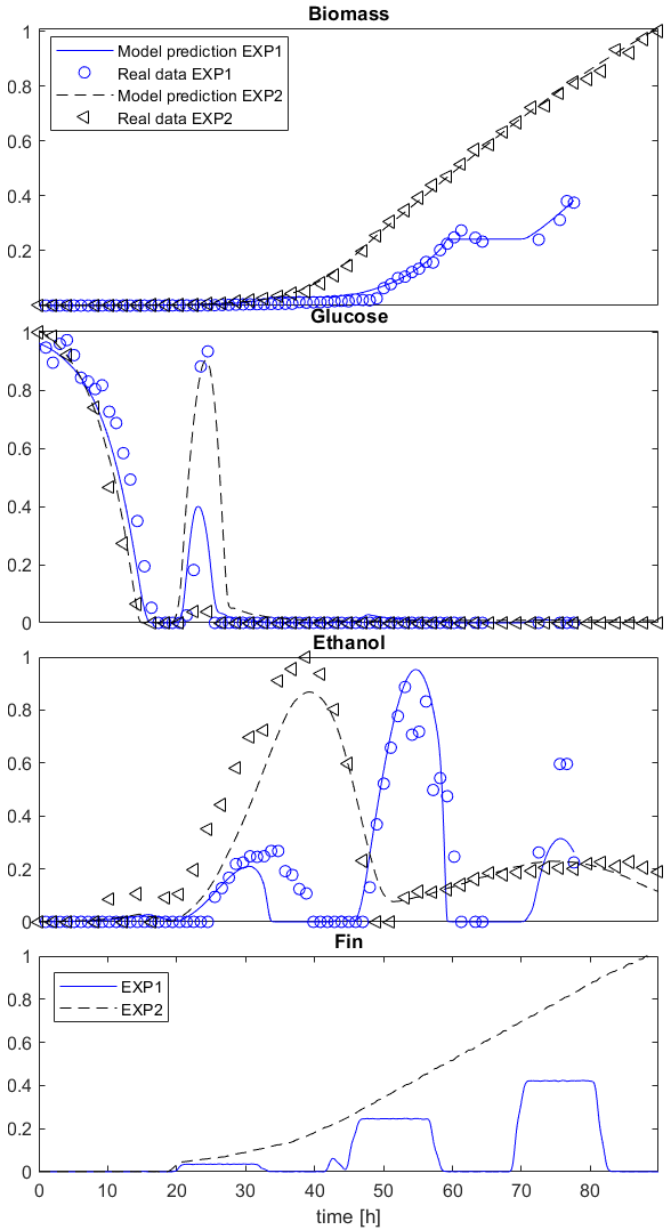

Fig. 1. Model direct validations using the two available experimental data sets

The model state predictions, for the 2 experiments, are plotted in Fig. 1.

3. NONLINEAR MODEL PREDICTIVE CONTROL

The objective of the control system is to maximize the production of biomass by regulating the fed-batch culture input flow. In Dewasme et al. (2015), different objective functions were proposed to optimize the biomass productivity of a bioreactor operated in fed-batch mode. One of the straightforward criteria considers the optimization of the biomass concentration X over the considered horizon. This criterion is represented by the following equation:

$$\Phi(t_k) = \sum_{i=1}^{p} X_{k+i} \tag{4}$$

which represents the volumetric productivity over a time horizon p and is maximized provided an adequate computation of the corresponding input F_{in} at time t_k. Moreover, restrictions are added to the control changes to avoid large variations of F_{in}. Hence, the augmented cost function is expressed as follows:

$$\Psi(t_k) = -\Phi(t_k) + \lambda \sum_{j=1}^{m} (F_{in,k+j-1} - F_{in,k+j-1}^{ref})^2 \tag{5}$$

where m is the control horizon with $m < p$, F_{in} is the manipulated variable, F_{in}^{ref} is the feed rate prediction reference obtained from the previous optimization and λ is the control penalty weight constant that is used to balance the two terms of the equation 5. It should be noticed that (5) is written in a form suggesting a minimization.

The NMPC is implemented considering the nonlinear model of equation 3 that can be formulated in a more generic form as:

$$\dot{z} = f(z, v, \theta) \tag{6}$$

where z is the state vector, v is the scalar input F_{in} and θ is the parameter vector. The NMPC optimal problem formulation can be stated as:

$$\min_{v} \Psi(t_k) \tag{7a}$$

$$\textbf{subject to} \quad \dot{z} = f(z(t), v(t), \theta), t \in [t_k, t_{k+p}] \tag{7b}$$

$$\boldsymbol{v}_{\mathrm{L}} \leq \boldsymbol{v}(t) \leq \boldsymbol{v}_{\mathrm{U}} \tag{7c}$$

$$\Delta v_L \leq \Delta v_{j-1} \leq \Delta v_V, \ j = 1, ..., m \tag{7d}$$

where the subscripts $_L$ and $_V$ stand respectively for lower and upper and $\boldsymbol{v} = \{F_{in,k}, ..., F_{in,k+m-1}\}$ is the feed rate policy over a control horizon of m sampling time intervals. Δv_{j-1} is the variation of the control action and it has m predicted values. The constraint given by equation (7d) enforces control moves, over the control horizon, to remain in a specific interval.

The simulation shown Figure 2 is run with the application of the NMPC in closed-loop, and uses the nominal model with parameters listed in table 1. The prediction and control horizons are respectively set to p=5 and m=2. A culture time of 90 hours is considered, as well as a sampling time of 6 minutes, and input limits of $0 \leq F_{in} \leq 0.3\,\mathrm{L\,h^{-1}}$. The lower and upper input bounds are $\Delta v_L = -0.03\,\mathrm{L\,h^{-1}}$ and $\Delta v_U = 0.03\,\mathrm{L\,h^{-1}}$, while λ is set at 10^4.

To maintain data confidentiality, Figure 2 magnitudes have been normalized based on the maximum values. The input vector over the time horizon m is updated at each sampling time following the numerical resolution of the NMPC problem (7). The latter is solved using the constrained nonlinear programming solver *fmincon* from the Matlab platform.

It can be observed from Figure 2 that the system was able to achieve a maximum biomass concentration value greater (approximately 1.5 times) than the experimental open-loop one while keeping low ethanol concentrations. However, even if these first results are promising, model uncertainties should be taken into account since they could affect the controller performance, providing undesired state trajectories. Parametric uncertainties are computed following the application of Gaussian variations determining 25 sets of parameters. The corresponding zero-mean distribution spreads the values with a standard deviation equal to the CVs from table 1.

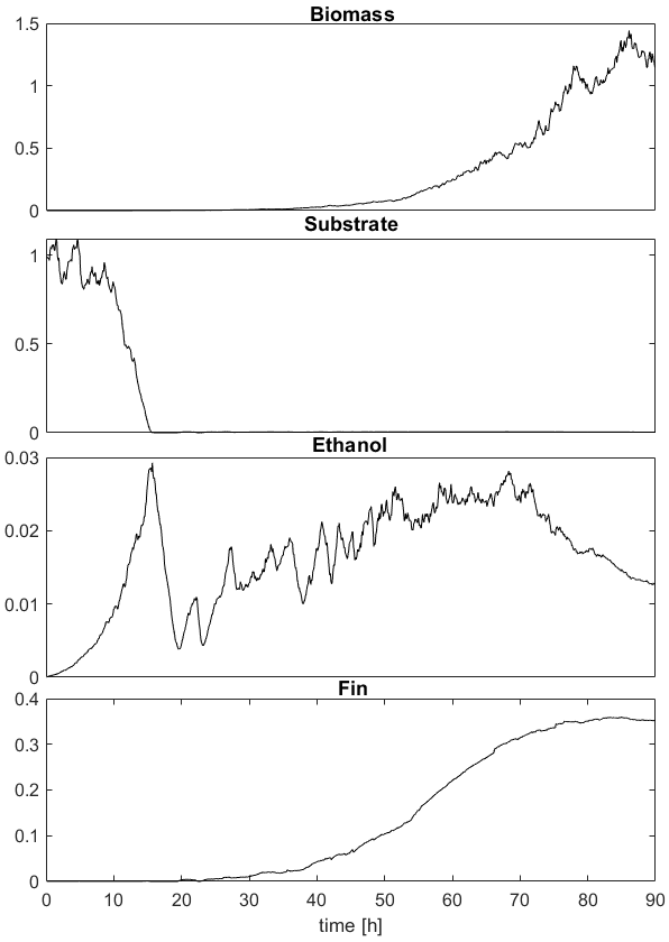

Fig. 2. NMPC simulation using a model with nominal parameters

Figure 3 shows the corresponding Monte-Carlo analysis, where the classical NMPC optimizes the biomass productivity during 25 runs considering the variations of the parameters K_{S2}, μ_{m2}, and $k2$, originating from the most uncertain reaction according to the CVs from table 1. For this new study, λ was set to 3 10^5. As highlighted in Figure 3, the simulations reveal a significant spreading of all input and output trajectories. Reducing these variations is a convenient target which is often required by quality-by-design procedures commonly applied in industrial manufacturing. A robust NMPC formulation is therefore necessary, aiming at guaranteeing that the trajectory spreading remains in a limited corridor.

4. TUBE-BASED NMPC

Tube-based NMPC allows tracking an assumed optimal nominal trajectory (i.e. using the nominal model), for instance, delivered by the classical NMPC structure solving (7), and considering an online adaptation of the input and state trajectories using a second ancillary controller which aims at minimizing the distance between the plant and the nominal trajectories. This strategy is illustrated in the block diagram presented in Figure 4.

For the sake of clarity, we recall the definition of the nominal model (6), which is assumed to be disturbance-free, conversely to the actual plant which is assumed to be

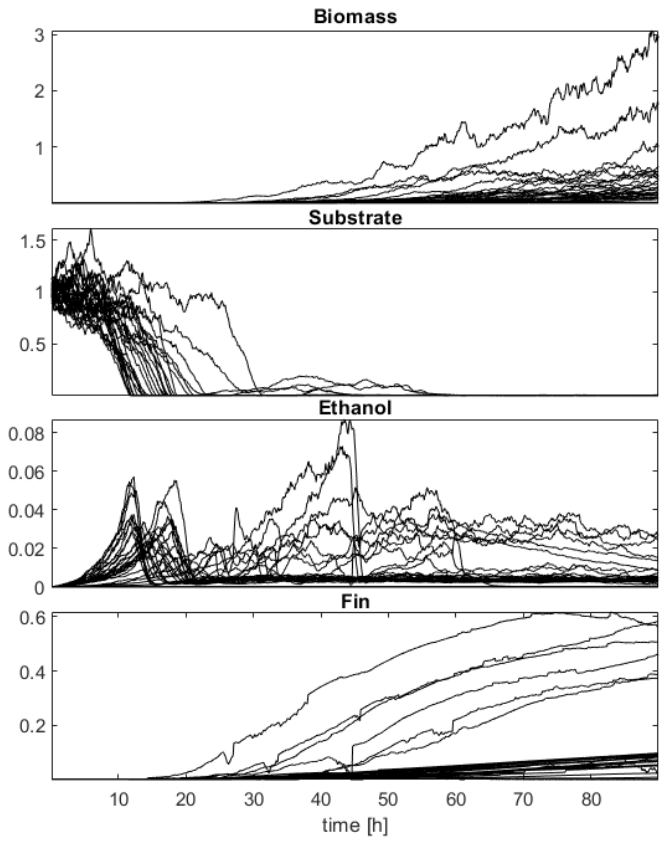

Fig. 3. Monte-Carlo analysis of the classical NMPC application during 30 runs where parameters K_{S2}, μ_{m2}, and $k2$ present a Gaussian distribution centered on the values from table 1 with their respective CV as standard deviation.

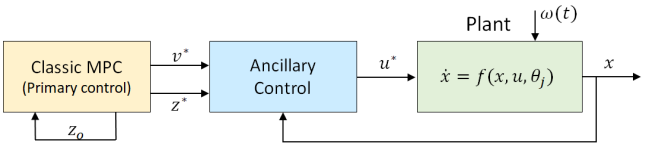

Fig. 4. Block diagram of the Tube-based NMPC implementation.

corrupted by a bounded model disturbance w, yielding:

$$\dot{z} = f(z, v, \theta) \tag{8a}$$
$$\dot{x} = f(x, u, \theta + w) \tag{8b}$$

where x and u are respectively the state and input vectors of the disturbed system, i.e. affected by the parameter uncertainties w. The nominal optimal state and input trajectories obtained by solving (7) are respectively denoted by $z^* = [X_z^* \ S_z^* \ E_z^*]$, and v^*.

The distance between these nominal optimal trajectories and the disturbed system state predictions can be formulated as follows:

$$\Phi_2(t_k) = a_1 \sum_{i=1}^{p} (X_{x|k+i} - X^*_{z|k+i})^2 +$$
$$a_2 \sum_{i=1}^{p} (S_{x|k+i} - S^*_{z|k+i})^2 + a_3 \sum_{i=1}^{p} (E_{x|k+i} - E^*_{z|k+i})^2 \tag{9}$$

where (X_x, S_x, E_x) are the disturbed system predictions calculated from the plant measurements at instant k, and the optimal states of the nominal model are (X_z^*, S_z^*, E_z^*). All state distances, considered as penalties, are assumed to be weighted, respectively, by a_1, a_2 and a_3.

The nonlinear programming problem of the ancillary NMPC is formulated as follows:

$$\min_u \Phi_2(t_k) + \lambda_2 \sum_{l=1}^{m} (u_{k+l-1} - v^*_{k+l-1})^2 \quad (10a)$$

$$\textbf{s.t.} \quad \dot{x} = f(x(t), u(t), \theta + w), t \in [t_k, t_{k+p}] \quad (10b)$$

$$\dot{z}^* = f(z^*(t), v^*(t), \theta), t \in [t_k, t_{k+p}] \quad (10c)$$

$$\boldsymbol{u}_{\mathrm{L}} \leq \boldsymbol{u} \leq \boldsymbol{u}_{\mathrm{U}} \quad (10d)$$

$$\Delta u_L \leq \Delta u_{j-1} \leq \Delta u_U, \; j = 1, ..., m \quad (10e)$$

where λ_2 is the weight applied to the input penalty. The initial condition of the nominal model used in the primary controller (7) is denoted $z_{0,k}$ and is updated at each time step $t = k$. The initial condition of the disturbed model is assumed to be the plant measurement available in k, denoted $x_{0,k}$. The values of a_1, a_2, and a_3 are set on the inverse of the maximum state values obtained during the model validation experiments. This allows the normalization of each penalty term and an equivalent spreading of their importance.

Figure 5 presents the results of the proposed tube-based NMPC strategy in a Monte-Carlo analysis where the same seed of random parametric deviations generated during the classical NMPC applications is considered, corrupting K_{S2}, μ_{m2}, and $k2$. The same state initialization is also applied and the controller parameterization is reported in table 2 as "Case 1". Regarding the nominal controller, the value of λ from the classical case is conserved ($\lambda = 3\ 10^5$). The trajectory tightening of the robust NMPC can be observed for all states and the input, improving the reproducibility of the operation despite the large parametric uncertainties. Reducing the trajectory spreading, however, imposes a tradeoff with the economic objective, and maximum biomass concentration levels attained in some of the classical NMPC runs may be higher than in the robust case. The biomass production reaches, on average, a normalized level of 0.408 in the classical NMPC case while the tube provides 0.302. However, the corresponding standard deviations are 0.59 for the classical case and 0.2 for the tube NMPC, validating its better reproducibility of the operating conditions.

Table 2. Tube-NMPC parameterizations during the Monte-Carlo analyzes.

Parameters	*Case* 1	*Case* 2
λ	$3\ 10^5$	10^4
λ_2	10^6	10^6
$a1$	0.02	0.02
$a2$	0.03	0.03
$a3$	20	20

Parameter tuning plays an important role in the tube NMPC behavior. To highlight this statement, a new Monte-Carlo analysis is run with a second parameterization also reported in table 2 as "Case 2", where λ is set to its initial value used in figure 2, i.e. 10^4. Reducing λ decreases the input variation penalty from (5) and favors biomass production. The new results are shown in figure 6, where the state corridors are larger while the input corridor is tighter than in case 1. Also, higher biomass productions are reached with an average final normalized biomass concentration of 1.08 with a standard deviation of 0.42 therefore showing a lower reproducibility of the operating conditions.

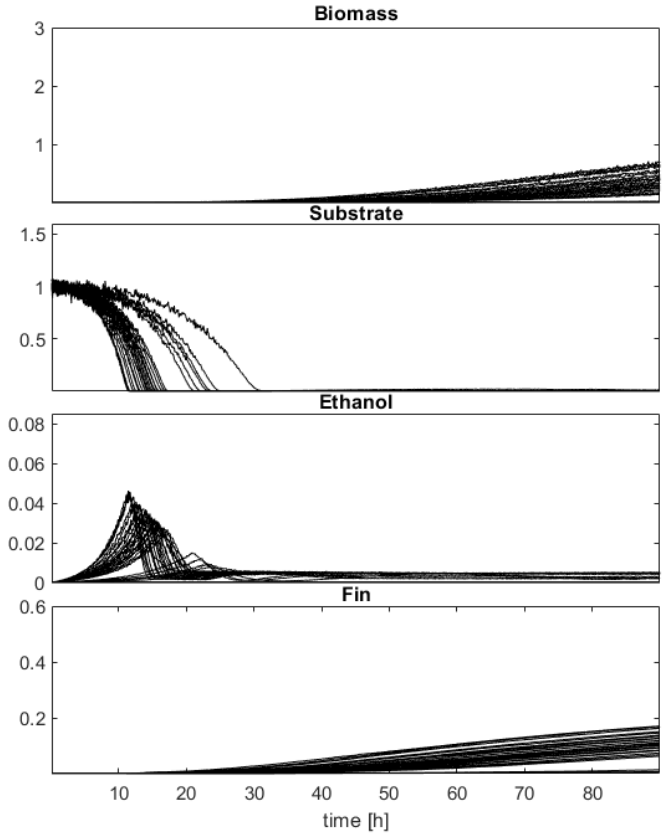

Fig. 5. Monte-Carlo analysis of the tube-based NMPC application during 25 runs where parameters K_{S2}, μ_{m2}, and $k2$ present a Gaussian distribution centered on the values from table 1 with their respective CV as standard deviation.

5. CONCLUSION

This paper discusses both classical and robust Nonlinear Model Predictive Control (NMPC) techniques for optimizing the production of biomass in yeast-based recombinant protein production. The controllers use a mechanistic model whose parameters are estimated using two experiment data sets. The resulting parameter estimates exhibit large coefficients of variations (CV) conferring an important uncertainty level to the model.

A Monte-Carlo analysis is carried out over 25 simulated runs, where the most uncertain reaction parameters are allowed to vary within their CV ranges. The results of the classical NMPC application show a significant dispersion in the state and input trajectories. This dispersion can be attributed to the high sensitivity of the model to the varying parameters.

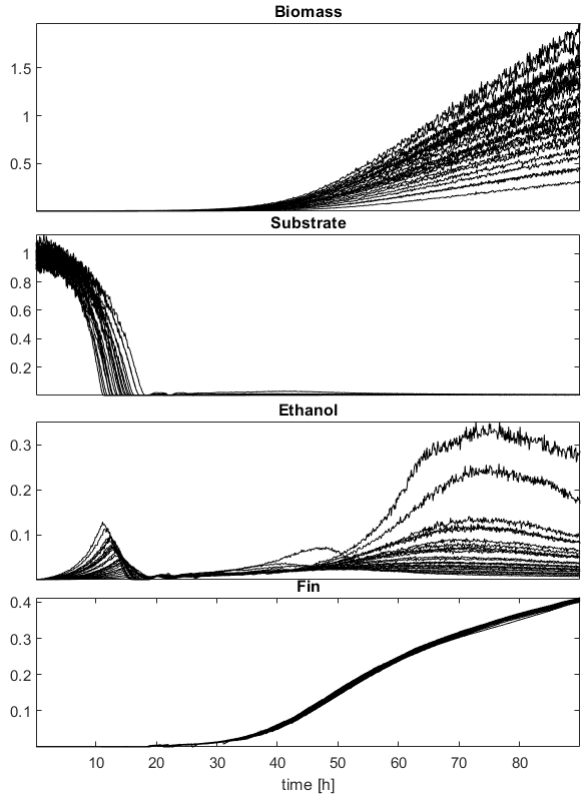

Fig. 6. Monte-Carlo analysis of the tube-based NMPC application during 25 runs where parameters K_{S2}, μ_{m2}, and $k2$ present a Gaussian distribution centered on the values from table 1 with their respective CV as standard deviation. Second parameterization.

The tube-based NMPC demonstrates improvements regarding the reproduction of the operating conditions and, therefore, the tightening of the state and input trajectories, despite the effect of parametric uncertainties. Eventually, the effect of the controller parameterization is highlighted by decreasing the input variation penalty in the nominal NMPC. The robustness paradigm is illustrated by the resulting tradeoff between high biomass production levels versus state and input trajectory tightening.

To reduce parameter uncertainty and limit the productivity/robustness tradeoff, additional experiments should be achieved. An economic assessment should complete the study to quantify the impact of the robust method on productivity.

ACKNOWLEDGEMENTS

The European researchers were funded by the Service Public de Wallonie (SPW), Belgium, under BElgian WAllonia REsearch (BEWARE) fellowships and European Union (EU) framework program for research and innovation, Marie Skłodowska-Curie Actions (MSCA).

CONFLICT OF INTEREST

P. Gerkens and G. de Lannoy are employees of the GSK group of companies.

REFERENCES

Akesson, M., Hagander, P., and Axelsson, J.P. (2001). Avoiding acetate accumulation in *Escherichia coli* cultures using feedback control of glucose feeding. *Biotechnology and Bioengineering*, 73(3), 223–230.

Angeli, D., Amrit, R., and Rawlings, J.B. (2012). On average performance and stability of economic model predictive control. *IEEE Transactions on Automatic Control*, 57(7), 1615–1626.

Crabtree, H. (1929). Observations on the carbohydrate metabolism of tumors. *Biochemical Journal*, 23, 536–545.

Dewasme, L., Fernandes, S., Amribt, Z., Santos, L., Bogaerts, P., and Vande Wouwer, A. (2015). State estimation and predictive control of fed-batch cultures of hybridoma cells. *Journal of Process Control*, 30, 50–57.

Dewasme, L., Mäkinen, M., and Chotteau, V. (2023). Practical data-driven modeling and robust predictive control of mammalian cell fed-batch process. *Computers & Chemical Engineering*, 171, 108164.

Dewasme, L., Mäkinen, M., and Chotteau, V. (2024). Multivariable robust tube-based nonlinear model predictive control of mammalian cell cultures. *Computers & Chemical Engineering*, 108592.

Dewasme, L., Srinivasan, B., Perrier, M., and Wouwer, A.V. (2011a). Extremum-seeking algorithm design for fed-batch cultures of microorganisms with overflow metabolism. *Journal of Process Control*, 21(7), 1092–1104.

Dewasme, L., Coutinho, D., and Wouwer, A.V. (2011b). Adaptive and robust linearizing control strategies for fed-batch cultures of microorganisms exhibiting overflow metabolism. In *Informatics in Control, Automation and Robotics: Revised and Selected Papers from the International Conference on Informatics in Control, Automation and Robotics 2010*, 283–305. Springer.

Hebing, L., Tran, F., Brandt, H., and Engell, S. (2020). Robust optimizing control of fermentation processes based on a set of structurally different process models. *Industrial and Engineering Chemistry Research*, 59(6), 2566–2580.

Qin, S.J. and Badgwell, T.A. (2000). An overview of nonlinear model predictive control applications. In F. Allgöwer and A. Zheng (eds.), *Nonlinear Model Predictive Control*. Birkhauser, Switzerland.

Qin, S.J. and Badgwell, T.A. (2003). A survey of industrial model predictive control technology. *Control Engineering Practice*, 11, 733 – 764.

Santos, L.O., Dewasme, L., Coutinho, D., and Vande Wouwer, A. (2012). Nonlinear model predictive control of fed-batch cultures of micro-organisms exhibiting overflow metabolism: Assessment and robustness. *Computers and Chemical Engineering*, 39, 143–151.

Silva, A.J.D., Rocha, C.K.d.S., and de Freitas, A.C. (2022). Standardization and key aspects of the development of whole yeast cell vaccines. *Pharmaceutics*, 14(12), 2792.

Sonnleitner, B. and Käppeli, O. (1986). Growth of saccharomyces cerevisiae is controlled by its limited respiratory capacity: formulation and verification of a hypothesis. *Biotechnology and bioengineering*, 28(6), 927–937.