Adv Biochem Eng Biotechnol
https://doi.org/10.1007/10_2020_149

Usage of Digital Twins Along a Typical Process Development Cycle

Peter Sinner, Sven Daume, Christoph Herwig, and Julian Kager

Contents

1 Introduction
2 Identification of Strain and Process Characteristics
3 Model-Based Process Design
4 Process Transfer and Model Lifecycle Management
5 Real-Time Usage of Digital Twins
5.1 Monitoring
5.2 Control
6 Conclusion
References

Abstract Digital methods for process design, monitoring, and control can convert classical trial-and-error bioprocess development to a quantitative engineering approach. By interconnecting hardware, software, data, and humans currently untapped process optimization potential can be accessed. The key component within such a framework is a digital twin interacting with its physical process counterpart. In this chapter, we show how digital twin guided process development can be applied on an exemplary microbial cultivation process. The usage of digital twins is described along a typical process development cycle, ranging from early strain characterization to real-time control applications. Along an illustrative case study on microbial upstream bioprocessing, we emphasize that digital twins can integrate

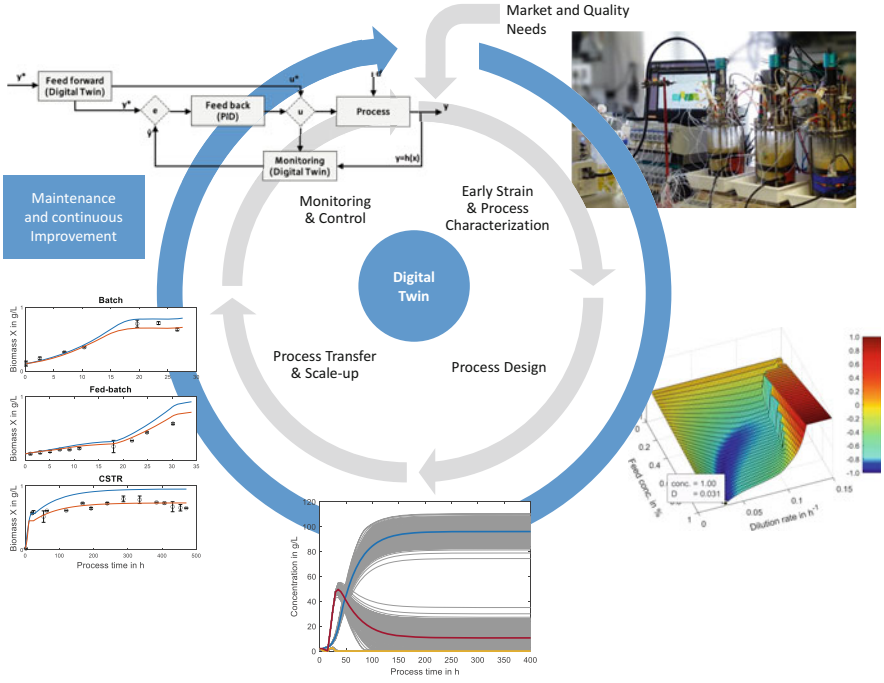
P. Sinner, S. Daume, and J. Kager (✉)
Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, Vienna, Austria
e-mail: julian.kager@tuwien.ac.at

C. Herwig
Institute of Chemical, Environmental and Bioscience Engineering, Technische Universität Wien, Vienna, Austria

Competence Center CHASE GmbH, Linz, Austria

entire process development cycles if the digital twin itself and the underlying models are continuously adapted to newly available data. Therefore, the digital twin can be regarded as a powerful knowledge management tool and a decision support system for efficient process development. Its full potential can be deployed in a real-time environment where targeted control actions can further improve process performance.

Graphical Abstract



Keywords Bioprocess development, Control, Digital twin, Dynamic modeling, Process systems engineering, Real-time monitoring

1 Introduction

Digital transformation gives analog devices and manual work a digital footprint. Spanning from the supply chain through the manufacturing process to the final products this digital footprint offers a wide range of initiatives toward a more competitive, flexible, and sustainable industry. Nevertheless, this transformation to computer-integrated manufacturing is only crowned with success when the symbiosis between hardware devices, software algorithms, and humans is highly

intensified. An essential step hereby is a virtual or digital representation of the physical asset, the so-called Digital Twin [1].

In terms of manufacturing the digital twin consists of a virtual representation of a production system, which is synchronized with all process-relevant data sources of the real system and able to run simulations, predictions, optimizations, and other actions based on different mathematical models and algorithms. This integration and synchronization with the real system and its specific application turns digital objects (a process model with manual data flow) into a digital twin [1]. Based on the current wave of digitalization, entire socio-technical structures are changed by transforming decision-making on various levels and creating new business models. Hereby, digital innovation is characterized by the re-programmable nature of digital objects, the homogenization of different data sources as well as positive network externalities [2]. Considering examples such as Amazon [3], digitalization can be regarded as a "disruptive technology" [4]. As digital objects are more and more evolving to fully integrated digital twins [5, 6], digitalization already substantially impacts and has the potential to further transform manufacturing as well.

Applications of digital twins in the field of manufacturing can be beneficial throughout the whole process development chain, see Fig. 1. During early process development simple mechanistic models [7] can be used for data evaluation and to plan first experimental designs including potential critical process parameters [8, 9]. New process knowledge can then be added to the basic model by either data-driven or if possible mechanistic terms. After testing and including all potentially influencing parameters into the model, the model can be seen as an digital twin

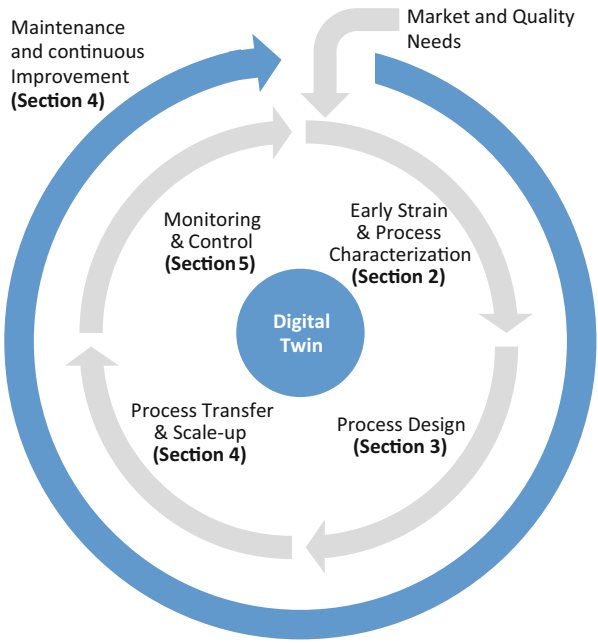


Fig. 1 Typical development cycle for a (pharmaceutical) bioprocess, with a digital twin as a central supportive element

and used for experimental design [8], process optimization [9] and thereafter for process monitoring [10, 11] and control on the running process [12, 13]. By using the digital twin throughout the process development chain, development times and costs can be reduced by replacing trial-and-error approaches by targeted in silico optimized experiments. Design space identification can contribute to a more comprehensive understanding of constraints in complex process systems. Additional, capacity potential can be quantified and harnessed, leading to higher process efficiencies. Aforementioned monitoring and control applications enable agile production practices as well as improved reproducibility and process reliability by an efficient response to process deviations. Especially for industries, facing increasing competitive pressure, e.g. in the fields of biosimilar manufacturing or biorefining, digital twin based methods are essential to ensure optimal operation of plants and thereby increasing their profitability.

Based on an exemplary upstream bioprocess this Chapter will elucidate some application examples of a digital twin throughout a typical process development cycle. As indicated in Fig. 1 the review includes digital twin usage for early strain and process characterization (Sect. 2), process design (Sect. 3), process transfer (Sect. 4), monitoring (Sect. 5.1), and control (Sect. 5.2) as well as maintenance and continuous improvement (Sect. 4).

2 Identification of Strain and Process Characteristics

Bioprocess engineers turn exciting discoveries into products and industrial processes. During early strain screening promising candidates are extracted toward maximum product titers and efficient utilization of specific substrates. Based on lab-automation [14, 15], miniaturization [16], and high parallelization [17, 18], a high number of strain variants can be generated and screened. Although these advances lead to a tremendous speed-up in strain engineering and selection processes, there is still minor quantitative knowledge on the selected strains, except its potentially advantageous behavior compared to the other candidates. Based on these simple preliminary screening experiments first important strain and process characteristics can be extracted and used for a more process-centered strain selection [9].

During digital twin development engineers can rely on an extensive repertoire of model classes. One of the most common classification schemes groups the process models regarding their extent of knowledge of the underlying mechanisms (“white,” “grey,” and “black box” models) [19]. Mechanistic models (“white box”) are based on the conservation of mass, energy, and momentum complemented by constitutive relations (e.g., reaction rate expressions). The establishment of mechanistic models requires a detailed prior knowledge of the underlying process phenomena and is in contrast to the development of data-driven models (“black box”) often exhausting and time-consuming [20]. The validity range of data-driven models lies only in the immediate neighborhood of the collected data and therefore lacks in extrapolation capabilities [21]. This is the core issue of data-driven models regarding their

applicability in bioprocess industry, where the information collected during a process is very scarce. These drawbacks can be partly overcome by a model class with increasing popularity, the so-called hybrid semi-parametric models ("gray box") [20].

Starting with some basic knowledge and hypotheses on the examined organism, it is possible to draft a first reaction scheme. Unstructured schemes, assuming a simplified metabolism, represent a good starting point as they only need the quantification of main reactants such as glucose, biomass, and product. The preliminary data is hereby scanned to detect biological phenomena, such as limitation, saturation, and inhibition [22]. At the cost of higher model complexity, structured schemes can be applied to additionally account for intracellular effects that can be relevant for strain characterization, see also Sect. 4. To achieve this, database information on metabolic networks [23, 24] and experimental omics data need to be available and integrated [25]. Further advances in high throughput omics analysis are critical to successfully include systems biology approaches into digital twin based strain characterization and design in the future [26, 27]. In the case study presented in this section, we focus on simple phenotype screening experiments readily accessible in any microbiology laboratory and show how such preliminary data can be viably used for model-based methods.

Within Fig. 2 a preliminary data set of a microbial strain growing on different substrates is given. Hereby, three shake flask experiments with two different carbon sources ($S1$ and $S2$) were conducted and biomass and residual sugar concentrations were measured. The poorly time-resolved information (only 5 measurement points) hinders data evaluation such as direct calculation of growth and conversion coefficients as well as smoothing and data-driven analysis. A visual inspection leads to the following hypothesized reaction scheme:

$$S1 \xrightarrow{q_{S1}} X$$

$$S2 \xrightarrow{q_{S2}} X$$

The organism can grow on both substrates with the specific rates q_{S1} and q_{S2} but prefers $S1$, which can be seen in the experiment where both sugars are present. For the two experiments with the single sugars it can be seen that $S1$ is consumed faster than $S2$. This hypothesized reaction scheme can be transformed to a set of ordinary differential equations (ODEs) shown in Eq. 1, which describes the mass balances of main components of biological systems including nonlinear and interacting reaction kinetics. In contrast to data-driven models, which aim to describe input–output correlations by transforming multidimensional input data sets to obtain maximal output accuracy, mechanistic models aim to describe the causal relationships within the process. Therefore, in addition to give reliable results, these models are an excellent and condensed summary of process knowledge.

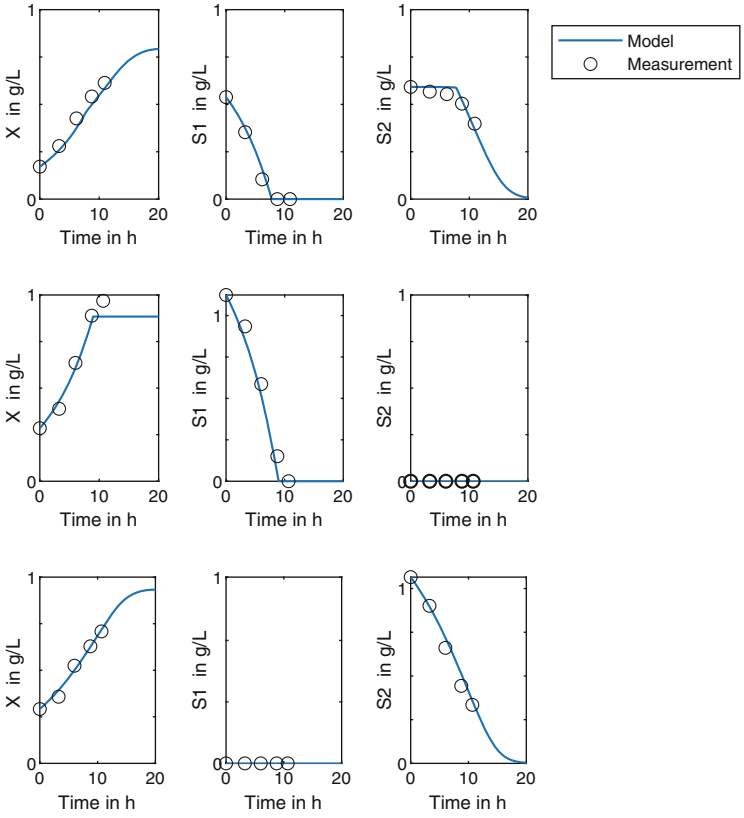


Fig. 2 Shake flask experiments of recombinant *C. glutamicum* on CGXII medium supplemented with a sugar mixture. Relative representations of biomass (X) and two carbon substrate concentrations (*S*1 and *S*2) are shown. Based on a simple kinetic model the data can be evaluated and strain characteristics, displayed in Table 1, can be deduced

$$\begin{aligned} \frac{dc_X}{dt} &= \mu c_X \\ \frac{dc_{S1}}{dt} &= -q_{S1} c_X \\ \frac{dc_{S2}}{dt} &= -q_{S2} c_X \end{aligned} \tag{1}$$

The ODE system describes the changes in concentrations of biomass (c_X), *S*1 (c_{S1}) and *S*2 (c_{S2}) over time in a batch reactor (shake flask in our case). The specific growth rate μ is a superposition of the two specific consumption rates and is coupled to these via the yield coefficients $Y_{X/S1}$ and $Y_{X/S2}$.

$$\mu = q_{S1} Y_{X/S1} + q_{S2} Y_{X/S2} \tag{2}$$

Monod kinetics for both specific substrate uptake rates were chosen

$$q_{S1} = q_{S1,max} \frac{c_{S1}}{c_{S1} + K_{S1}} \tag{3}$$

$$q_{S2} = q_{S2,max} \frac{c_{S2}}{c_{S2} + K_{S2,app}} \tag{4}$$

with the maximum specific substrate uptake rate $q_{S1,\ max}$ and the affinity constant K_{S1} for $S1$ and analogously with $q_{S2,\ max}$ and $K_{S2,\ app}$ for $S2$.

In the first shake flask experiment with both sugars available, it was observed that $S2$ was nearly not taken up until $S1$ was fully consumed. Therefore, it was hypothesized that a competitive inhibition must occur (Eq. 5). This is a good example for the flexibility of mechanistic models. When new insights are gained into the process this model type can be extended at will. To account for the inhibition effect an additional term for the saturation constant was incorporated into the consumption kinetic of $S2$. In the presence of $S1$ the saturation constant $K_{S2,\ app}$ ($K_{S2} >> K_{S1}$) is linearly increased.

$$K_{S2,app} = K_{S2} + \frac{K_{S2}}{K_{S1}} c_{S1} \tag{5}$$

The hypothesized model can be tested on the available shake flask data, by fitting the model to the data as described in [7]. Hereby, the model parameters given in Table 1 are changed to minimize the residuals between model simulations and measurements. Although the shake flasks can be evaluated individually, their results are biased by erroneous measurements and incomplete data. Therefore, it is preferable to simultaneously fit the three data sets, which results in the final parameters with their associated errors, which are displayed in Table 1.

In order to assess model quality, goodness of fit as well as parametric uncertainty are determined and compared to predefined acceptance criteria. The acceptance criteria are dependent on the targeted application of the model and on the importance of the modeled states. Evaluation of model quality is necessary to choose the appropriate model, e.g. out of a library of various model structures, and to decide if additional experimental data is required. Goodness of fit measures such as the root-

Table 1 Identified model parameters from a simultaneous fit with their uncertainties, calculated by the inverse of the overall Fisher information further described in [7]

Parameter	Value	Unit	Rel. Error in %
$q_{S1,\ max}$	0.24	g/(gh)	0.53
$Y_{X/S1}$	0.58	g/g	3.1
$q_{S2,\ max}$	0.21	g/(gh)	0.52
$Y_{X/S2}$	0.57	g/g	4.1

mean-square error (RMSE) or the normalized root-mean-square error (NRMSE) are commonly used. Parametric uncertainty can be derived from the inverse of the Fisher information matrix [7] and stated as relative parameter error in %, also referred to as coefficient of variation (CV). An NRMSE below 15% for the targeted states and parametric uncertainty below 40% [28, 29] are defined as acceptance criteria in this case study. Overall, the hypothesized model describing the growth on two substrates matches well with the underlying shake flask data sets (NRMSE below 5.1%, 6.5%, 14.1% for biomass, $S1$, and $S2$ states, respectively) and the resulting parameters of maximum uptake rates and yield coefficients show reliable values with low errors (CV below 5%). This first, identified model builds a solid basis for further usage in process development.

Already at the stage of early process development and strain characterization, when little data availability hampers the usage of statistical evaluation procedures and the direct calculation of conclusive strain characteristics, mechanistic mass balance description can be of great value. Different hypotheses can be tested and valuable information can be retrieved by a simultaneous evaluation of all available data sets, although they are subject to measurement errors and have a poor temporal resolution.

3 Model-Based Process Design

To facilitate the determination of the process design space, an in silico investigation of the system behavior based on preliminary models can complement classical Design of Experiment (DoE) approaches [30]. DoE approaches enable to describe the design space with a reduced number of experiments, but a DoE is of very limited information if the included factors and their selected levels are not appropriate [31]. Because of their extrapolation capabilities mechanistic models are indispensable for in silico investigation of process performance [21].

As displayed in Fig. 3 the design space can be seen as the overlaying region between technical and biological possibilities leading to the required product and process outcome. Based on the relatively simple process models, as presented in

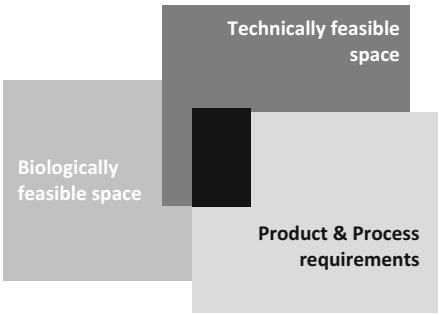


Fig. 3 Schematic process design space presented as the overlaying regions between biological and technical constraints and the process and product requirements

Sect. 2, the identified biological behavior can be combined with technical constraints and analyzed, without the need of any experiment. This digital exploration is of great value in process development and enables to conduct a suitable DoE including the most influential factors and correct deflection levels. Besides that, potential biological and technical bottlenecks can be identified.

For the exemplary microbial strain it could be required to design a process to convert a sugar stream, containing substrate 1 and substrate 2 with a fixed relation ($c_{S2,\,in} \approx 3 \cdot c_{S1,\,in}$), into biomass. The process aims to convert the sugars efficiently to biomass in a continuous manner, which can be expressed as an objective J normalized between -1 (maximal biomass formation rate) and 1 (no sugar consumed).

$$J = -D \cdot c_X + c_{S1} + c_{S2} \tag{6}$$

To the batch model given in Eq. 1 an input feed F_{in} with substrate concentration c_Si,in is added as well as the dilution rate D defined as $\frac{F_i n}{V}$ to enable the simulation of continuous cultures. To additionally include the oxygen transfer as a technical constraint the model can be extended by the dissolved oxygen concentration c_{O_2} by including the reactor specific oxygen transfer ($k_L a = 300\,\mathrm{h}^{-1}$), the oxygen solubility ($c^*_{O_2} = 0.2088\,\mathrm{mmol/L}$), and the oxygen to biomass conversion coefficient ($Y_{O_2/X}$), determined by elemental balance assuming purely oxidative growth.

$$\begin{aligned}
\frac{dc_X}{dt} &= (\mu - D)\, c_X \\
\frac{dc_{S1}}{dt} &= -q_{S1}\, c_X + D(c_{S1,in} - c_{S1}) \\
\frac{dc_{S2}}{dt} &= -q_{S2}\, c_X + D(c_{S2,in} - c_{S2}) \\
\frac{dc_{O_2}}{dt} &= -\mu\, Y_{O_2/X} + k_L a \left(c^*_{O_2} - c_{O_2}\right)
\end{aligned} \tag{7}$$

The identified substrate uptake kinetics given in Eqs. 3 and 4 were extended by an additional Monod term to include oxygen as a limiting component, with an affinity constant K_{O_2}.

$$q_S = f(c_S)\, \frac{c_{O_2}}{c_{O_2} + K_{O_2}} \tag{8}$$

Based on this set of differential equations, model simulations of continuous processes (CSTR) with different dilution rates and sugar concentrations in the inflow can be calculated and the respective objective function can be evaluated. Within Fig. 4 the resulting digital design space of the given process system is displayed. For combinations of dilution rates from 0.03 to 0.1 h^{-1} and feed concentrations from 50 to 100% best performing processes can be identified. The theoretical optimum (minimum) is reached with a dilution rate of 0.031 h^{-1} and a feed concentration of 100%.

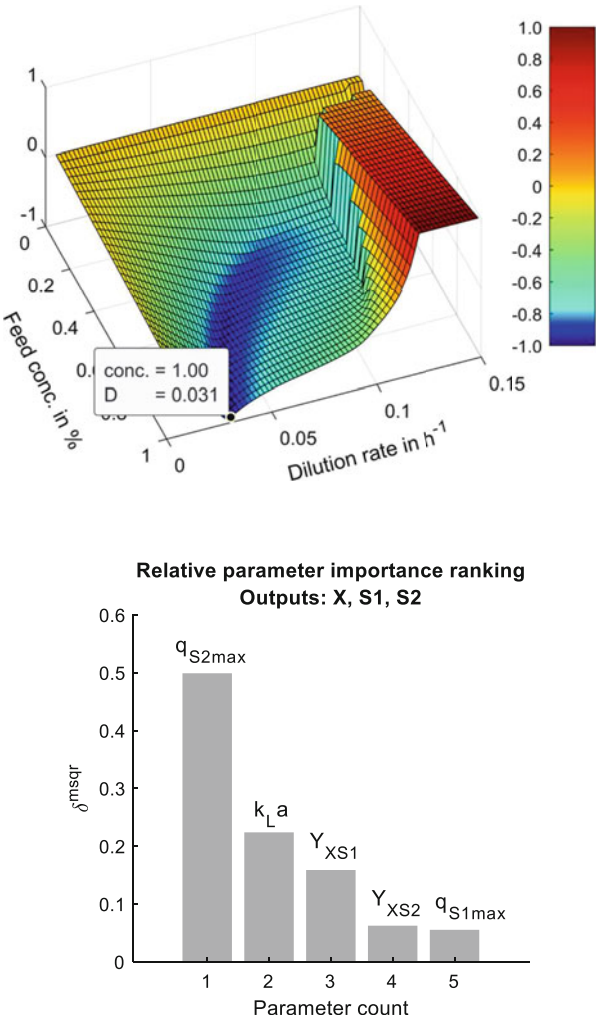


Fig. 4 Top: Digital process design space in function of dilution rate and feed concentration including biological growth behavior and technical limitations to obtain best performing processes
Bottom: Relative parameter senstitvities evaluated at the indicated optimum to reveal potential bottlenecks

By local parametric sensitivity analysis a deeper insight on the predicted optimum can be gained. The local model sensitivities were calculated according to the procedure described in [7] and the overall impact of model parameters on the three main outputs, biomass (c_X) and substrates (c_{S1} and c_{S2}), is displayed as the relative root-mean-square error (msqr) in Fig. 4. The sensitivity analysis reveals that at the optimal operating point the process is mostly influenced by the maximum uptake rate of $S2$ (q_{S2}) and the reactor specific maximum volumetric oxygen mass transfer coefficient (k_La). To further improve the process performance, higher oxygen transfer of the reactor as well as strain variants with a better uptake of $S2$ would be necessary.

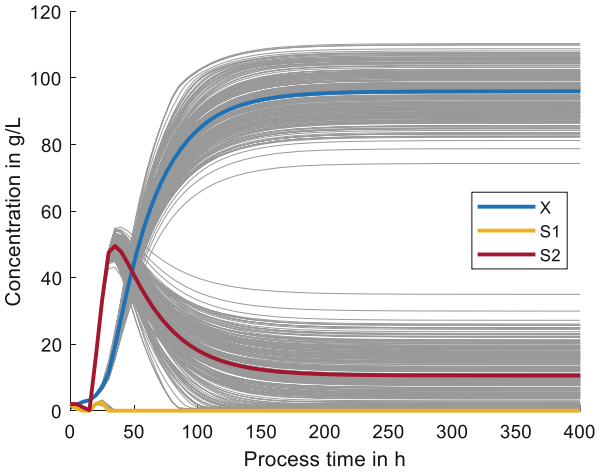


Fig. 5 Uncertainty analysis of a continuous process at its optimum operating point. The colored lines indicate the original output and the disturbed outputs out of 200 simulations are shown as gray lines. Gaussian sampling from the parameter errors given in Table 1 was conducted

To foresee the stability of the optimal operation point as well as to analyze the robustness of the analyzed process, Monte Carlo simulations are of great value [32, 33]. Hereby, simulations are repeated by disturbing the model parameters within their expected error range [34], given in Table 1. As displayed in Fig. 5 we can see that the relatively small parameter errors (below 5%) show a significant impact on the different outputs. The colored lines indicate the original output and the disturbed outputs out of 200 simulations are given in grey. Whereas $S1$ is only slightly effected by the perturbations and reaches limiting concentrations in all simulations, the biomass and $S2$ concentrations are highly affected. This indicates that the real process can possibly reach biomass concentrations between 80 and 110 g/L and $S2$ concentrations in the outflow between 0 and 30 g/L.

Although the main process outcomes are highly uncertain, the process reached steady state conditions in every simulation. Therefore the determined optimum can be considered as a feasible operating point. Besides that, it is notable that the process is subject to a very long startup phase of approx. 200 h until steady state conditions are reached. Based on model simulations of other similarly performing operating points and startup conditions the startup phase can be possibly reduced or at least be foreseen for a potential process. Further insights into the system response and stability can be derived by model analysis methods, using the eigenvalues of the Jacobian matrix [35].

4 Process Transfer and Model Lifecycle Management

An initial digital twin was developed in Sect. 2 and applied for the (digital) design of a CSTR operation in Sect. 3. In order to transfer initial screening experiments to the actual bioreactor conditions, targeted experiments need to be performed. New data sets generated during process transfer are continuously integrated throughout the

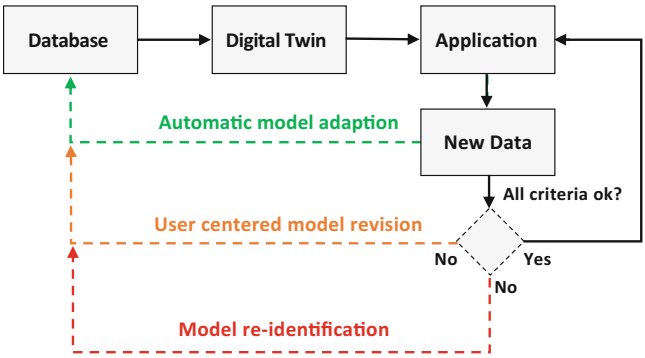


Fig. 6 Model lifecycle management along a digital twin guided bioprocess development cycle

concomitant life cycling of the digital twin and the corresponding process under development. Therefore, model lifecycle management is needed to update the digital twin as outlined in Fig. 6.

A dynamic model is a valuable tool to plan experiments during process development. As described in Sect. 3 the design space describing key performance parameters of the target process, e.g. space-time yields, can be assessed by model-based methods. Based on an identified design space operational conditions can be selected to approach the actual experimental optimum in a targeted way [36, 37].

Albeit only initial knowledge about the system might be available at an early stage, promising operational conditions such as substrate feeding trajectories and dilution rates can be estimated by the digital twin and are then validated experimentally. In our exemplary process different operational modes (batch, fed-batch, and continuous) were applied in bioreactors to generate further data after shake flask screening and to approach the real application scenario of an up-scaled process. This aims to identify discrepancies between shake flask results and different bioreactor operational modes and to illustrate how a digital twin can be adapted accordingly. In this regard, also scale-down bioreactor approaches can be applied to simulate large-scale effects already during process development [38, 39]. Using the digital twin developed in Sects. 2 and 3 technically feasible and physiologically relevant experimental conditions were selected for the three different operational modes. Model-based methods can also be applied to design experimental plans and sampling schedules that directly optimize the information output to further improve the parametrization of process models [8, 40–42]. The obtained results after process transfer are shown in Fig. 7. In silico designed trajectories that were simulated using the prototype model available from shake flask screening can qualitatively describe growth in bioreactor conditions. However, only batch bioreactor behavior was predicted with acceptable accuracy. The newly available data needs to be integrated to improve the applicability of the digital twin. Therefore, the underlying process models need to be continuously adapted to account for new insights, gained during bioprocess development, see Fig. 6. In addition to new experimental data this can also include changing technical limitations or performance targets as well as access

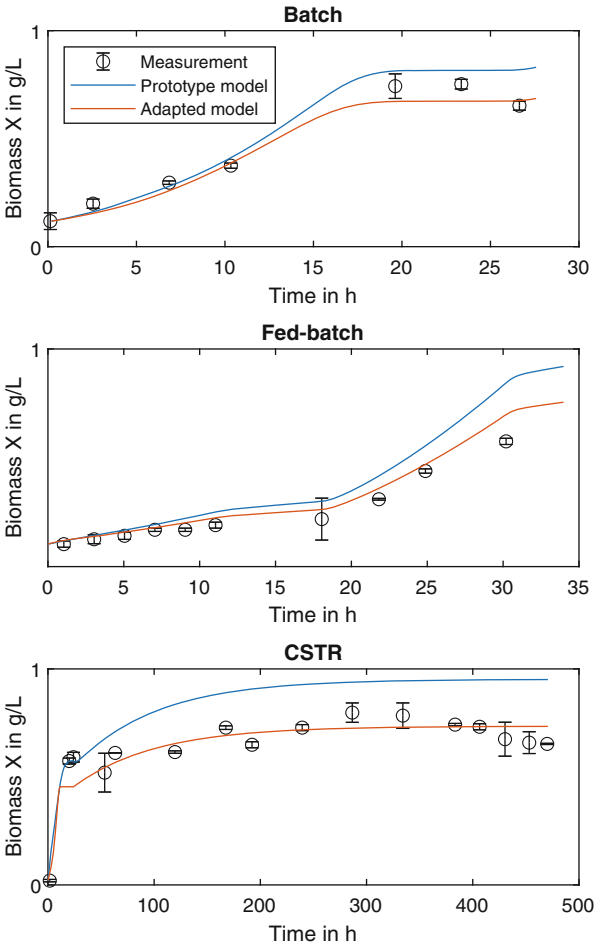


Fig. 7 Process transfer from shake flask screening (prototype model) to batch, fed-batch, and continuous (CSTR) bioreactor cultivations of recombinant *C. glutamicum* on CGXII medium supplemented with a sugar mixture. Biomass concentrations (X) are shown as relative representations. The process model is automatically adapted by integrating newly available data sets in order to improve model validity and transferability

to new strain variants. Knowledge management is hereby highly relevant to successfully implement QbD principles [43] from screening to manufacturing. Target-oriented workflows for data analysis and model generation [7, 44] are necessary to realize this concept in practice and proceed toward automated evolution of digital twins [45]. A digital twin in use should not be considered as "finished" but as an evolving tool that needs maintenance [46] and enables continuous process improvement.

In the first process transfer and model update step shown in Fig. 7 automatic model adaption by parameter re-estimation was sufficient to extend the digital twin to fed-batch and CSTR applications. Using multiple data sets in combination can considerably improve parameter identifiability and model transferability. When additional bioreactor data sets were combined within one parametrization step, parameter errors were reduced by a factor of 10 or more, as can be seen in

Table 2 Model parameter uncertainties obtained by combination of multiple data sets: shake flask screening (initial) and further bioreactor experiments (updated)

Parameter	Rel. Error in %	
	(Initial)	(Updated)
$q_{S1,\ max}$	0.53	0.013
$Y_{X/S1}$	3.1	0.13
$q_{S2,\ max}$	0.52	0.0027
$Y_{X/S2}$	4.1	0.039

Table 2. However, automatic parameter adaption can not account for any new phenomena that are currently not contained in the model structure. If these effects are critical to reach the predefined process goals, the model structure has to be revised [7]. Therefore, new kinetic descriptions are added, e.g. by extending Monod equations with different limiting substrates or inhibitory compounds. For an overview on widely used kinetic descriptions, the reader is referred to [7, 44]. In addition to changing interactions between already implemented states also new states can be introduced, e.g. to account for by-product effects that were not considered in the previous prototype model structure. Subsequently, each model revision step is validated by analyzing the acceptance criteria for model quality, such as NRMSE on target states, and parameter identifiability [44]. In the case study presented here, the addition of a death rate to account for late batch-phase biomass decline, see Fig. 7, is a possibility to more accurately describe the process dynamics if conditions of biomass decline are relevant for the target operational mode.

Here, we started with a simple unstructured and unsegregated model. In case this approach fails to describe process behavior, intracellular metabolism and population heterogeneity can be included to account, e.g. for metabolic transitions caused by scale-up effects. The effect of intracellular reactions on bioprocess behavior can be described by metabolic flux analysis [47, 48], also considering methods for dynamic flux balance analysis [49]. Recent approaches also include regulatory elements in the network structure [50, 51]. To deal with the high uncertainty of complex biochemical networks, ensemble modeling techniques can be applied [52, 53]. Besides intracellular effects, also population heterogeneity can be integrated into a digital twin if it is needed to describe critical process behavior. Models can be extended by Euler-Lagrange CFD methods to describe the trajectories of individual cells, and thereby the population heterogeneity in reactors [54, 55]. The combination of intracellular metabolism and CFD simulations, yielding structured segregated models, holds potential for a more targeted design of large-scale bioprocesses [56].

When increasing model complexity by additional equations or parameters there is often a trade-off between accuracy on training data and model transferability. Cross-validation with data sets from different operational modes can help to prevent overfitting. This is particularly challenging when purely data-driven methods are used and only few and poorly time-resolved data sets are available. As this is typically the case during early process development we recommend to include mechanistic elements such as mass balances to facilitate process transfer and robust real-time usage.

5 Real-Time Usage of Digital Twins

A multitude of further process optimization opportunities can be made accessible when the interaction between process setup and digital twin is transferred to a real-time setting. Process disturbances and deviations can then be monitored (Sect. 5.1) during running operations and control actions (Sect. 5.2) can be taken in a timely and targeted manner to prevent process failure. To achieve this goal a process environment enabling bidirectional data transfer between plant devices and the digital twin is needed. Nowadays, bioprocess engineers have a variety of digitization tools at hand to implement real-time functionality for any laboratory, pilot, or production facility. Data transfer via open platform communications (OPC) technologies enables interconnection of hardware devices, sensors, and numerical computing interfaces such as Python or Matlab, as displayed in Fig. 8 for a lab-scale setup. The numerical computing interface serves as a flexible platform for observer and controller design. Resulting digital twin elements that are ready for application can then be transcribed to programmable logic controllers (PLC) used in the respective manufacturing environment. Additionally, different industrial process information management systems (PIMS) are available and provide a human–machine interface further simplifying bioprocess digitalization.

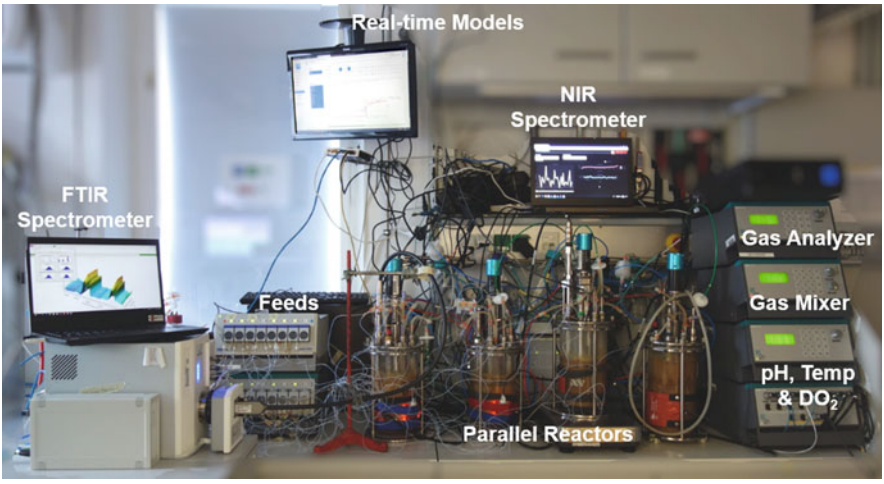


Fig. 8 Digitized bioreactor setup at TU Wien (Institute of Chemical, Environmental and Bioscience Engineering). Interconnection of hardware (sensors and actuators) and software components enables digital twin based monitoring and control in real-time

5.1 Monitoring

Real-time information is of crucial importance to evaluate the culture state during bioprocess operations and to adapt process parameters accordingly. With model-based approaches unmeasured states, such as biomass concentrations, or states that are not measurable in real-time, such as biomass specific rates, can be estimated continuously. To achieve this goal available primary measurements, e.g. substrate concentrations measured with a time delay at-line, and secondary measurements, e.g. online offgas analysis or in situ spectroscopy such as FTIR or NIR, need to be linked to known process inputs such as substrate feed rates and previous process knowledge in the form of mathematical models. This constitutes the next phase of digital twin lifecycling as the digital twin is now adapted for real-time usage. As a first step the process model structure needs to be extended for the relevant measurements used for state estimation or soft sensing. The resulting measurement model functions, see Eq. 10, link the current system states to the available measurements. The dynamics of system states (e.g., biomass and carbon substrate concentrations) are described by the system model functions Eq. 9 containing the ODE system developed previously in Eq. 7. The selection of information-rich measurements can be facilitated by observability analysis [57, 58]. For our example process, outlined in the previous sections, carbon evolution (CER) and oxygen uptake rates (OUR) calculated from online offgas analysis [59] were selected as secondary measurements to estimate system states during a fed-batch bioprocess. Offgas dynamics can be described by first principle carbon and redox balances [60]. Alternatively, empirical yield coefficients linking offgas dynamics and growth behavior can be estimated based on experimental data and introduced as additional model parameters [61, 62].

$$\dot{\underline{x}} = f(t, \underline{x}, \underline{u}, \underline{\theta}) = (\dot{c}_X, \dot{c}_{Si}) \tag{9}$$

$$\underline{y} = h(t, \underline{x}, \underline{u}, \underline{\theta}) = (CER, OUR) \tag{10}$$

Real-time monitoring strategies using process models in combination with hardware sensors can be regarded as soft(ware) sensors [63]. A variety of soft sensor strategies are available for up- and downstream processing, such as data-driven [64, 65], kinetic [11, 66], elemental balancing [60], or hybrid [67–69] modeling approaches. It is highly recommended to include filtering algorithms for state estimation to account for process and measurement noise as well as model uncertainty. As bioprocesses show nonlinear behaviors, nonlinear filters such as the extended [70] and unscented [71] Kalman filter or the particle filter [72] are commonly applied. Here, we show how particle filtering can be used for real-time estimation of the most probable biomass concentration in our exemplary fed-batch process, see Fig. 9. As an comparison, the prototype model generated during shake flask screening, see Sect. 2, is applied as an model simulation without any real-time information and as an digital twin including real-time information of CER and OUR. Whereas regular simulations with the prototype model can only inaccurately

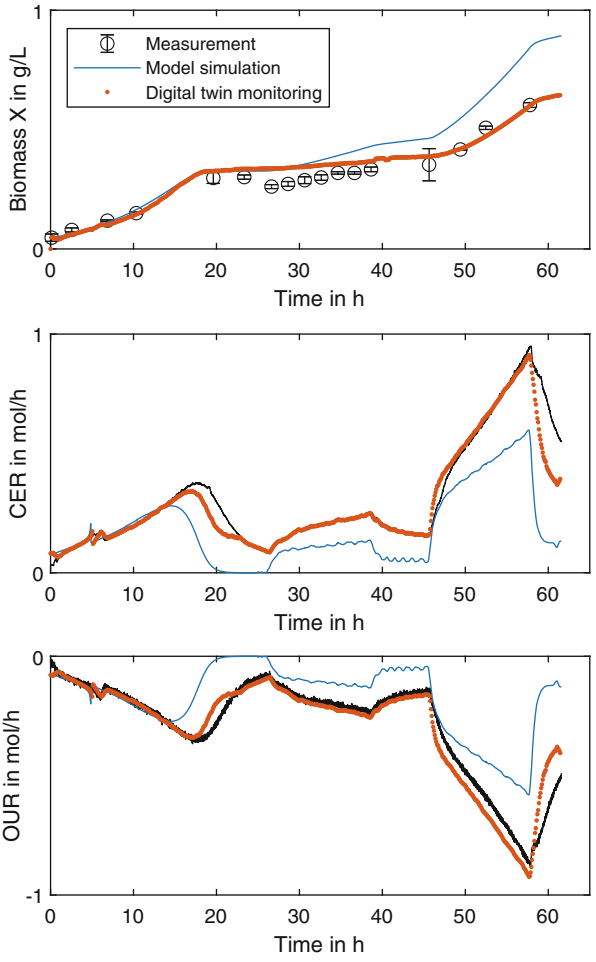


Fig. 9 Digital twin based real-time monitoring of biomass during a fed-batch bioprocess of recombinant *C. glutamicum* on CGXII medium supplemented with a sugar mixture. Biomass concentrations (X), carbon evolution rates (CER), and oxygen uptake rates (OUR) are shown as relative representations. A prototype model from shake flask screening is integrated in a particle filtering framework for state estimation using offgas data (CER, OUR) as secondary measurements

describe fed-batch behavior, its use within a state estimator framework yields a satisfactory monitoring accuracy.

It is important to consider the reliability of the measured data before using it as input for the soft sensor. Ideally, this is done in a process-phase dependent or time-resolved manner. Including additional sensor equipment can considerably improve the monitoring performance. But the respective added information should be carefully evaluated to avoid an unnecessary increase of setup complexity that can limit practical applicability, e.g. in a bio-pharmaceutical manufacturing environment. Model-plant mismatch, caused, for example, by changes in lag phase behavior during cultivation startup or deviations resulting from batch to fed-batch process transfer, can be partly corrected by a state estimation algorithm during running processes. However, if critical process parameters are not covered in the system and measurement models, these effects need to be further studied experimentally and

incorporated in the digital twin used for real-time monitoring, see model lifecycle management in Sect. 4. Once reliable digital twin based monitoring is available, it can be further deployed for real-time process control as outlined in the following section.

5.2 Control

Within the final step, process control strategies need to be defined and implemented in the developed process [73]. Today, a broad range of control strategies are used for biochemical processes, which reach from the control of physico-chemical entities within the reactors up to sterilization and cleaning in place, feeding strategies, harvesting as well as downstream processing by centrifugation, filtration, and chromatographic methods. Besides defining the optimal setpoints as shown in Sect. 3 it is important to maintain the setpoints during production and to define feasible transitions between setpoints or setpoint trajectories.

Deviations from the setpoint or harsh operational changes can lead to significant product loss, or in the worst case to a total failure of the process. Biochemical processes are very sensitive against changes in process parameters, so little overshoots in temperature and pH control as well as short oxygen or nutrient limitations can have irreversible effects on the organisms. This sensitiveness requires to act in a predictive manner as corrective measures based on standard feed-back controllers (e.g., step control or PID) can be too late or not effective to converge to the aimed setpoint. A digital twin can hereby help to predict the needed process actions in function of a predictive model, which can then be fine-tuned by a feed-back signal as displayed in Fig. 10. The feed-forward term of this so-called two-degree-of-freedom control [74] or also feed-forward, feed-back control [75] can consist of a predictive optimization model or a linearized model able to directly give the currently needed manipulated input variables u^*.

To obtain a direct calculation of the model output the model can be inverted by the so-called feed-back linearization. For the examined exemplary process this can be easily done for the uptake of q_{S1} for setpoints ($q_{S1_{\text{setpoint}}} < q_{S1,max}$) as the uptake kinetics are solely dependent on the related concentration of substrate 1 ($S1$). For

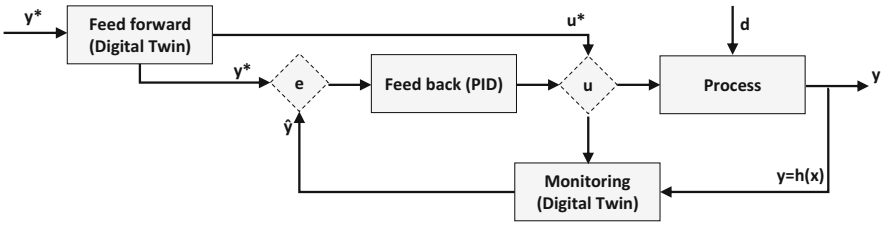


Fig. 10 Two degree of freedom control with digital twin based feed forward part and monitoring of the process output as described in Sect. 5.1

feed-back linearization of more complex and interacting systems mathematical procedures exist, which can be found in [76] and were successfully applied in biotechnological processes as shown in [13, 77, 78]. For the specific uptake rate q_{S1} the differential equation for substrate 1, given in Eq. 7, can be reformulated under steady state assumption $\frac{c_{S1}}{dt} = 0$:

$$F_{in} = \frac{V}{c_{S1,in} - c_{S1}} c_X q_{S1_{\text{setpoint}}} \tag{11}$$

to obtain the linearized control law with F_{in} being the manipulable variable u as a function of the setpoint (q_S1, setpoint) $q_{S1_{\text{setpoint}}}$, the current biomass c_X and the sugar concentration c_{S1} and the feed concentration $c_{S1,\ in}$.

Within Fig. 11 fed-batch simulations of different q_{S1} setpoints are shown. The first graph contains the feed profiles. After reaching a maximum volume of 3 L the processes were stopped. In the second graph the resulting q_{S1} profiles are displayed. After a relatively short transition phase all setpoints can be reached by the applied feed profiles. In dependence of the set q_{S1} different uptake rates q_{S2} can be observed. For the second substrate a much longer time to reach a stable uptake can be seen.

Based on the observed results a relation between q_{S1} and q_{S2} can be established, which is displayed in Fig. 12. Under usage of the feed with a fixed ratio of $S2 = 3 \cdot S1$ a q_{S1} between 0.06 and 0.2 leads to a stable q_{S2} of around 0.2. Between $q_{S1} = 0$ and 0.06 only little substrate 2 ($S2$) is accumulated, whereas with higher q_{S1} setpoints higher growth (μ) can be reached albeit at the cost of high accumulation of $S2$. Based on the objective to efficiently convert both substrates into biomass a q_{S1} setpoint of 0.06 seems to be the best trade-off, which is shown in Fig. 13.

Similarly as described in Sect. 3 a potential oxygen limitation can be additionally considered in the model. Hereby, the feed rate region where the oxygen concentration is below 30% of oxygen saturation is indicated with a gray box, in Fig. 13. We can conclude that the aimed setpoint can only be maintained around 15 h, before

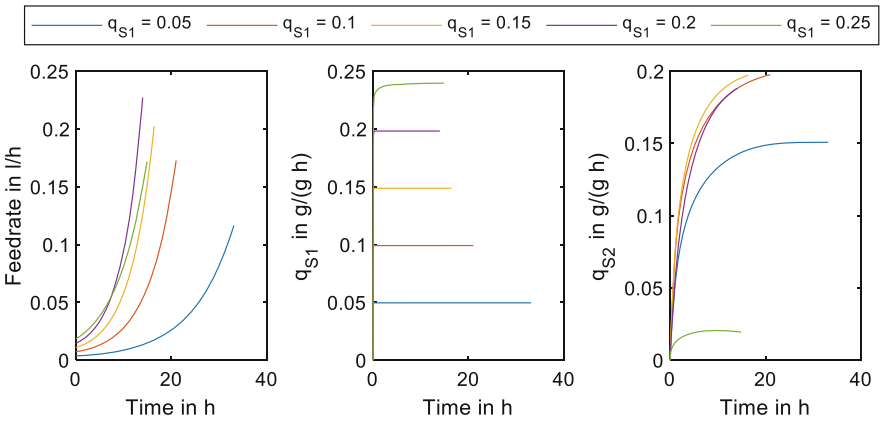


Fig. 11 Feed-forward control simulation with different q_{S1} setpoints and the resulting q_{S2} values

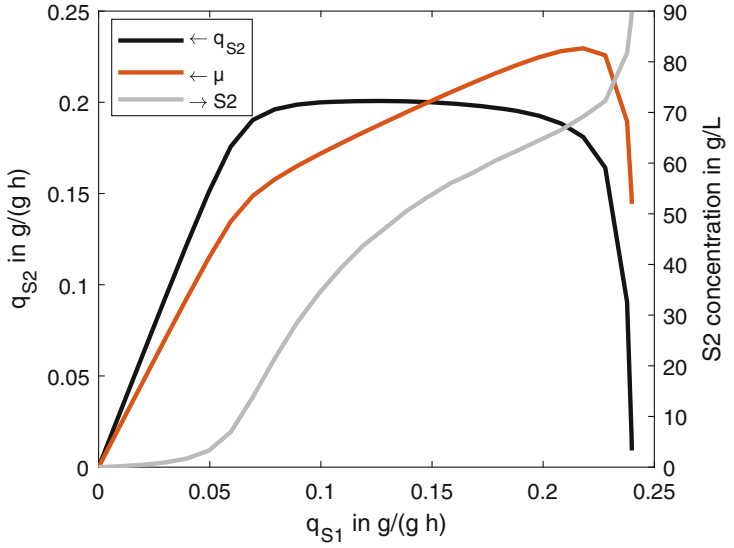


Fig. 12 Phase plot between applied q_{S1} setpoint and obtained q_{S2}, μ, and c_{S2} after reaching a volume of 3 L of simulated fed-batch experiments

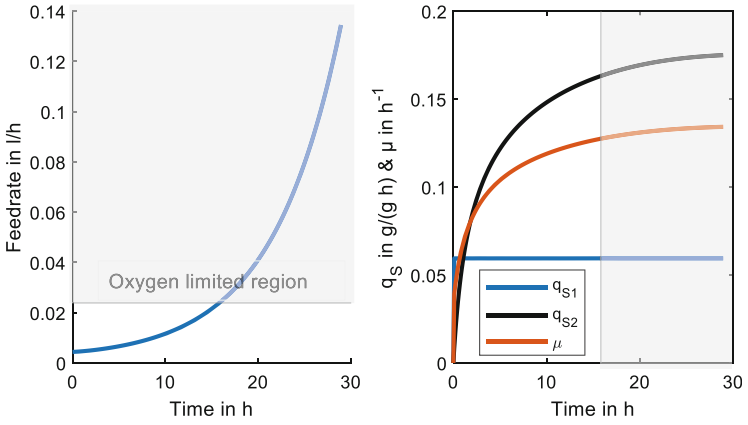


Fig. 13 Feed-forward control simulation of a q_{S1} of 0.06 with indicated regions of oxygen limitation with a k_La of 300 h^{-1} for a potential fed-batch process

growth is limited by the oxygen transfer. Therefore, an even lower setpoint ($q_{S1} = 0.02$) was selected to avoid oxygen limitation throughout the whole process duration.

Within Fig. 14 the results of a controlled reactor run are displayed. Based on the inverted model, the feed rate to obtain the aimed q_{S1} setpoint was applied on a running process. Based on the model the expected substrate 2 uptake q_{S2} could be predicted as well, which is in good accordance with the measured values. Also the

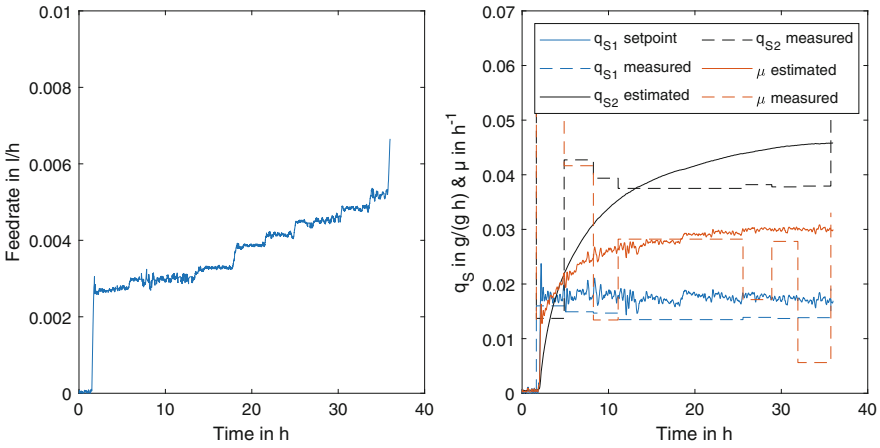


Fig. 14 Model-based control of biomass specific substrate uptake rate (q_{S1}) in a microbial fed-batch bioprocess utilizing a substrate mixture

predicted overall growth (μ) is in good accordance with the measurements. Within this simple example of feed-back linearization it could be shown how a digital twin can be used to control non-directly measurable entities, such as specific reaction rates, which is of great value for biotechnological processes.

To finally integrate digital twin based monitoring and control strategies into industrial operations their applicability needs to be validated. In the context of regulatory requirements, interpretability of algorithms is important to pinpoint cause and effect of control actions, which can be more readily achieved when using mechanistic or hybrid, mechanistic data-driven models. Furthermore, model-based robustness analysis is necessary to assess controller robustness against different process deviations, e.g. by using Monte Carlo simulations as described in Sect. 3. Subsequently, repeated experimental verification runs can ensure that process reliability requirements are met under target operational conditions. When process tolerance limits are not exceeded the validated monitoring and control strategy can be put in use.

6 Conclusion

The complexity of bioprocesses demands high efforts toward the full process development cycle. The integration of digital twins can help to deal with this complexity. Besides having a clear application goal governing the overall digital twin development process, a suitable software environment is key to transform the model into an integrated digital twin. In addition to software functionalities including structured model building and analysis workflows, data and model lifecycle management is important. Based on an existing framework, displayed within Fig. 6,

novel algorithms and use cases will be continuously added and tested for their applicability on biotechnological processes. With this solid basis the development of models and their transformation into functional digital twins is greatly enhanced. For the application of model-based solutions for monitoring and control, as described in Sect. 5 a real-time architecture is needed. This usually consists of a centralized database with the ability to directly communicate with the hardware devices and sensors and a numerical computing interface able to implement, maintain, and adapt the developed model-based solutions. Based on the OPC UA standard available today, this interconnection of database, software, and hardware devices was significantly simplified. Despite these technical solutions are accessible there are still different hurdles for the successful implementation of digital twins in today's industry. Firstly, development costs of model-based monitoring and control tools are currently high and retrofitting of existing installations can be challenging due to regulatory reasons, especially in the bio-pharmaceutical industry. Automated generation of model-based solutions and a more quantitative assessment of process optimization potential as well as associated risks can further promote digital twin usage. Another important driver is operator acceptance, as model-based decision-making that is not understandable to the personnel involved in plant control is unlikely to be supported. Human–machine interfaces that make digital twin decisions explainable, also under usage of machine learning approaches, are needed. Additionally, process systems engineering is only slowly becoming a more relevant part of the standard curriculum of biochemical engineering study programs. Once progress is made to overcome the aforementioned hurdles, computer-integrated manufacturing can be employed more efficiently to make Biotech industry ready for future challenges.

Acknowledgements This work has received funding from the Bio Based Industries Joint Undertaking under the European Union's Horizon 2020 research and innovation program under grant agreement No. 790507 (iFermenter) and was supported by the Austrian Research Promotion Agency (FFG) under the scope of the "ADAMO" project (No 864705). Further funding was provided by the Competence Center CHASE GmbH, funded by the Austrian Research Promotion Agency (FFG) (No 868615).

The authors would like to thank Marlene Stiegler for her assistance during experimental data generation.

References

1. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W (2018) Digital twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine 51(11):1016–1022
2. Yoo Y, Henfridsson O, Lyytinen K (2010) The new organizing logic of digital innovation: an agenda for information systems research. Inf Syst Res 21(4):724–735
3. Kimble C, Bourdon I (2013) The link among information technology, business models, and strategic breakthroughs: examples from Amazon, Dell, and eBay. Glob Bus Organ Excell 33 (1):58–68

4. Bower JL, Christensen CM (1995) Disruptive technologies: catching the wave. Harv Bus Rev 73(1):43–53
5. Lim KYH, Zheng P, Chen C-H (2019) A state-of-the-art survey of digital twin: techniques, engineering product lifecycle management and business innovation perspectives. J Intell Manuf:1–25
6. Negri E, Fumagalli L, Macchi M (2017) A review of the roles of digital twin in CPS-based production systems. Procedia Manuf 11:939–948
7. Daume S, Kofler S, Kager J, Kroll P, Herwig C (2020) Generic workflow for the setup of mechanistic process models. In: Animal cell biotechnology. Springer, Berlin, pp 189–211
8. Daume S, Kager J, Herwig C (2019) Time resolved sensitivity & identifiability analysis for directed parametrization of highly dynamic models. In: Computer aided chemical engineering, vol 46. Elsevier, Amsterdam, pp 1111–1116
9. Sinner P, Kager J, Daume S, Herwig C (2019) Model-based analysis and optimisation of a continuous *Corynebacterium glutamicum* bioprocess utilizing lignocellulosic waste. IFAC-PapersOnLine 52(26):181–186
10. Kager J, Berezhinskiy V, Zimmerleiter R, Brandstetter M, Herwig C (2019) Extension of a particle filter for bioprocess state estimation using invasive and non-invasive IR measurements. In: Computer aided chemical engineering, vol 46. Elsevier, Amsterdam, pp 1417–1422
11. Kager J, Herwig C, Stelzer IV (2018) State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. Chem Eng Sci 177:234–244
12. Ulonska S, Waldschitz D, Kager J, Herwig C (2018) Model predictive control in comparison to elemental balance control in an *E. coli* fed-batch. Chem Eng Sci 191:459–467
13. Kager J, Tuveri A, Ulonska S, Kroll P, Herwig C (2019) Experimental verification and comparison of model predictive, PID and model inversion control in a *Penicillium chrysogenum* fed-batch process. Process Biochem
14. Chao R, Mishra S, Si T, Zhao H (2017) Engineering biological systems using automated biofoundries. Metab Eng 42:98–108
15. Janzen NH, Striedner G, Jarmer J, Voigtmann M, Abad S, Reinisch D (2019) Implementation of a fully automated microbial cultivation platform for strain and process screening. Biotechnol J 14(10):1800625
16. Hemmerich J, Noack S, Wiechert W, Oldiges M (2018) Microbioreactor systems for accelerated bioprocess development. Biotechnol J 13(4):1700141
17. Wilming A, Bähr C, Kamerke C, Büchs J (2014) Fed-batch operation in special microtiter plates: a new method for screening under production conditions. J Ind Microbiol Biotechnol 41 (3):513–525
18. Kuivanen J, Holmström S, Lehtinen B, Penttilä M, Jäntti J (2018) A high-throughput workflow for CRISPR/Cas9 mediated combinatorial promoter replacements and phenotype characterization in yeast. Biotechnol J 13(9):1700593
19. Cameron IT, Hangos K (2001) Process modelling and model analysis. Elsevier, Amsterdam
20. Von Stosch M, Oliveira R, Peres J, de Azevedo SF (2014) Hybrid semi-parametric modeling in process systems engineering: past, present and future. Comput Chem Eng 60:86–101
21. Gernaey KV, Lantz AE, Tufvesson P, Woodley JM, Sin G (2010) Application of mechanistic models to fermentation and biocatalysis for next-generation processes. Trends Biotechnol 28 (7):346–354
22. Herold S, King R (2014) Automatic identification of structured process models based on biological phenomena detected in (fed-) batch experiments. Bioprocess Biosyst Eng 37 (7):1289–1304
23. Kanehisa M et al (2002) The KEGG database. In: Novartis foundation symposium. Wiley Online Library, pp 91–100
24. Caspi R, Billington R, Ferrer L, Foerster H, Fulcher CA, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA et al (2016) The MetaCyc database of metabolic

pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res 44(D1):D471–D480
25. Pinu FR, Beale DJ, Paten AM, Kouremenos K, Swarup S, Schirra HJ, Wishart D (2019) Systems biology and multi-omics integration: Viewpoints from the metabolomics research community. Metabolites 9(4):76
26. Marcellin E, Nielsen LK (2018) Advances in analytical tools for high throughput strain engineering. Curr Opin Biotechnol 54:33–40
27. Couvillion SP, Zhu Y, Nagy G, Adkins JN, Ansong C, Renslow RS, Piehowski PD, Ibrahim YM, Kelly RT, Metz TO (2019) New mass spectrometry technologies contributing towards comprehensive and high throughput omics analyses of single cells. Analyst 144(3):794–807
28. Ulonska S, Kroll P, Fricke J, Clemens C, Voges R, Müller MM, Herwig C (2018) Workflow for target-oriented parametrization of an enhanced mechanistic cell culture model. Biotechnol J 13 (4):1700395
29. Anane E, Barz T, Sin G, Gernaey KV, Neubauer P, Bournazou MNC et al (2019) Output uncertainty of dynamic growth models: effect of uncertain parameter estimates on model reliability. Biochem Eng J 150:107247
30. Garcia-Munoz S, Luciani CV, Vaidyaraman S, Seibert KD (2015) Definition of design spaces using mechanistic models and geometric projections of probability maps. Org Process Res Dev 19(8):1012–1023
31. Möller J, Kuchemüller KB, Steinmetz T, Koopmann KS, Pörtner R (2019) Model-assisted design of experiments as a concept for knowledge-based bioprocess development. Bioprocess Biosyst Eng 42(5):867–882
32. Kroese DP, Brereton T, Taimre T, Botev ZI (2014) Why the Monte Carlo method is so important today. WIREs Comput Stats 6(6):386–392
33. Biwer A, Griffith S, Cooney C (2005) Uncertainty analysis of penicillin V production using Monte Carlo simulation. Biotechnol Bioeng 90(2):167–179
34. Möller J, Rodrguez TH, Müller J, Arndt L, Kuchemüller KB, Frahm B, Eibl R, Eibl D, Pörtner R (2020) Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. Comput Chem Eng 134:106693
35. Kremling A (2013) Systems biology: mathematical modeling and model analysis. Chapman and Hall/CRC, Boca Raton
36. Kotidis P, Demis P, Goey CH, Correa E, McIntosh C, Trepekli S, Shah N, Klymenko OV, Kontoravdi C (2019) Constrained global sensitivity analysis for bioprocess design space identification. Comput Chem Eng 125:558–568
37. Weise T, Grewe C, Pfaff M (2020) Experimental and model-based analysis to optimize microalgal biomass productivity in a pilot-scale tubular photobioreactor. Front Bioeng Biotechnol 8:453
38. Neubauer P, Junne S (2010) Scale-down simulators for metabolic analysis of large-scale bioprocesses. Curr Opin Biotechnol 21(1):114–121
39. Anane E, García ÁC, Haby B, Hans S, Krausch N, Krewinkel M, Hauptmann P, Neubauer P, Bournazou MNC (2019) A model-based framework for parallel scale-down fed-batch cultivations in mini-bioreactors for accelerated phenotyping. Biotechnol Bioeng 116(11):2906–2918
40. Galvanin F, Macchietto S, Bezzo F (2007) Model-based design of parallel experiments. Ind Eng Chem Res 46(3):871–882
41. Bournazou MNC, Barz T, Nickel DB, Cárdenas DCL, Glauche F, Knepper A, Neubauer P (2017) Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. Biotechnol Bioeng 114(3):610–619
42. Abt V, Barz T, Cruz-Bournazou MN, Herwig C, Kroll P, Möller J, Pörtner R, Schenkendorf R (2018) Model-based tools for optimal experiments in bioprocess engineering. Curr Opin Chem Eng 22:244–252
43. Rathore AS, Garcia-Aponte OF, Golabgir A, Vallejo-Diaz BM, Herwig C (2017) Role of knowledge management in development and lifecycle management of biopharmaceuticals. Pharm Res 34(2):243–256

44. Kroll P, Hofer A, Stelzer IV, Herwig C (2017) Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering. Process Biochem 62:24–36
45. Hummer W, Muthusamy V, Rausch T, Dube P, El Maghraoui K, Murthi A, Oum P (2019) ModelOps: cloud-based lifecycle management for reliable and trusted AI. In: 2019 IEEE international conference on cloud engineering (IC2E). IEEE, pp 113–120
46. Flåten GR (2018) Model maintenance. In: Multivariate analysis in the pharmaceutical industry. Elsevier, Amsterdam, pp 313–321
47. Herwig C, von Stockar U (2002) A small metabolic flux model to identify transient metabolic regulations in *Saccharomyces cerevisiae*. Bioprocess Biosyst Eng 24(6):395–403
48. Hofer A, Hauer S, Kroll P, Herwig C (2018) Metabolic flux analysis linked to complex raw materials as tool for bioprocess improvement. Chem Eng Sci 191:245–252
49. Gomez JA, Barton PI (2018) Dynamic flux balance analysis using DFBAlab. In: Metabolic network reconstruction and modeling. Springer, pp 353–370
50. Shen F, Sun R, Yao J, Li J, Liu Q, Price ND, Liu C, Wang Z (2019) OptRAM: in-silico strain design via integrative regulatory-metabolic network modeling. PLoS Comput Biol 15(3): e1006835
51. Kwon MS, Lee BT, Lee SY, Kim HU (2020) Modeling regulatory networks using machine learning for systems metabolic engineering. Curr Opin Biotechnol 65:163–170
52. Kremling A, Geiselmann J, Ropers D, de Jong H (2018) An ensemble of mathematical models showing diauxic growth behaviour. BMC Syst Biol 12(1):1–16
53. Bromig L, Kremling A, Marin-Sanguino A (2020) Understanding biochemical design principles with ensembles of canonical non-linear models. PLoS One 15(4):e0230599
54. Haringa C, Noorman HJ, Mudde RF (2017) Lagrangian modeling of hydrodynamic–kinetic interactions in (bio) chemical reactors: practical implementation and setup guidelines. Chem Eng Sci 157:159–168
55. Kuschel M, Siebler F, Takors R (2017) Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. Bioengineering 4(2):27
56. Wang G, Haringa C, Tang W, Noorman H, Chu J, Zhuang Y, Zhang S (2020) Coupled metabolic-hydrodynamic modeling enabling rational scale-up of industrial bioprocesses. Biotechnol Bioeng 117(3):844–867
57. Golabgir A, Hoch T, Zhariy M, Herwig C (2015) Observability analysis of biochemical process models as a valuable tool for the development of mechanistic soft sensors. Biotechnol Prog 31 (6):1703–1715
58. Hermann R, Krener A (1977) Nonlinear controllability and observability. IEEE Trans Autom Control 22(5):728–740
59. Aehle M, Kuprijanov A, Schaepe S, Simutis R, Lübbert A (2011) Simplified off-gas analyses in animal cell cultures for process monitoring and control purposes. Biotechnol Lett 33(11):2103
60. Herwig C, Marison I, Von Stockar U (2001) On-line stoichiometry and identification of metabolic state under dynamic process conditions. Biotechnol Bioeng 75(3):345–354
61. Bastin G, Dochain D (1990) Dynamic models of bioreactors. In: Bastin G, Dochain D (eds) On-line estimation and adaptive control of bioreactors, process measurement and control. Elsevier, Amsterdam, pp 1–82
62. Van Impe JF, Bastin G, De Moor B, Van Breusegem V, Vandewalle J (1992) Optimal adaptive control of fed-batch fermentation process with growth/production decoupling. IFAC Proc Vol 25(2):351–354
63. De Assis AJ, Filho RM (2000) Soft sensors development for on-line bioreactor state estimation. Comput Chem Eng 24(2–7):1099–1103
64. Jie Y (2012) A Bayesian inference based two-stage support vector regression framework for soft sensor development in batch bioprocesses. Comput Chem Eng 41:134–144
65. Gopakumar V, Tiwari S, Rahman I (2018) A deep learning based data driven soft sensor for bioprocesses. Biochem Eng J 136:28–39
66. Natarajan S, Lee JH (2000) Repetitive model predictive control applied to a simulated moving bed chromatography system. Comput Chem Eng 24(2–7):1127–1133

67. Krämer D, King R (2019) A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. J Process Control 82:91–104
68. Feidl F, Garbellini S, Luna MF, Vogg S, Souquet J, Broly H, Morbidelli M, Butté A (2019) Combining mechanistic modeling and Raman spectroscopy for monitoring antibody chromatographic purification. Processes 7(10):683
69. Lopez PC, Udugama IA, Thomsen ST, Roslander C, Junicke H, Mauricio-Iglesias M, Gernaey KV (2020) Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation. Biofuels Bioprod Biorefin
70. Cox H (1964) On the estimation of state variables and parameters for noisy dynamic systems. IEEE Trans Autom Control 9(1):5–12
71. Julier SJ, Uhlmann JK (1997) New extension of the Kalman filter to nonlinear systems. In: Signal processing, sensor fusion, and target recognition VI, vol 3068. International Society for Optics and Photonics, pp 182–193
72. Gordon NJ, Salmond DJ, Smith AFM (1993) Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: IEE proceedings F (radar and signal processing), vol 140. IET, pp 107–113
73. Simutis R, Lübbert A (2015) Bioreactor control improves bioprocess performance. Biotechnol J 10(8):1115–1130
74. Hagenmeyer V, Nohr M (2008) Flatness-based two-degree-of-freedom control of industrial semi-batch reactors using a new observation model for an extended Kalman filter approach. Int J Control 81(3):428–438
75. Arndt M, Hitzmann B (2001) Feed forward/feedback control of glucose concentration during cultivation of *Escherichia coli*. IFAC Proc Vol 34(5):403–407
76. Isidori A (2013) Nonlinear control systems. Springer Science & Business Media
77. Mjalli FS, Al-Asheh S (2005) Neural-networks-based feedback linearization versus model predictive control of continuous alcoholic fermentation process. Chem Eng Technol 28 (10):1191–1200
78. Torres I, Queinnec I, Wouwer AV (2010) Observer-based output feedback linearizing control applied to a denitrification reactor. IFAC Proc Vol 43(6):102–107