Systems Microbiology and Biomanufacturing (2021) 1:257–274
https://doi.org/10.1007/s43393-021-00024-0

REVIEW

Digital Twin in biomanufacturing: challenges and opportunities towards its implementation

Isuru A. Udugama[1] · Pau C. Lopez[1] · Carina L. Gargalo[1] · Xueliang Li[2] · Christoph Bayer[3] · Krist V. Gernaey[1]

Received: 14 December 2020 / Revised: 1 February 2021 / Accepted: 3 February 2021 / Published online: 1 March 2021

Abstract
The domain of industrial biomanufacturing is enthusiastically embracing the concept of Digital Twin, owing to its promises of increased process efficiency and resource utilisation. However, Digital Twin in biomanufacturing is not yet clearly defined and this sector of the industry is falling behind the others in terms of its implementation. On the other hand, some of the benefits of Digital Twin seem to overlap with the more established practices of process control and optimization, and the term is vaguely used in different scenarios. In an attempt to clarify this issue, we investigate this overlap for the specific case of fermentation operation, a central step in many biomanufacturing processes. Based on this investigation, a framework built upon a five-step pathway starting from a basic steady-state process model is proposed to develop a fully-fledged Digital Twin. For demonstration purposes, the framework is applied to a bench-scale second-generation ethanol fermentation process as a case study. It is proposed that the success or failure of a fully-fledged Digital Twin implementation is determined by key factors that comprise the role of modelling, human operator actions, and other propositions of economic value.

Keywords Digital Twin · Fermentation · Process control · Modelling · Biomanufacturing

Introduction

Leveraging process data to derive meaningful insights, which, in turn, can be used to determine actions to improve plant performance, has been an area of interest for decades. The permeation of Industry 4.0 concepts to the process industries has made both the academia and the industry look into new technologies that can be employed to improve process efficiency, product quality as well as safety. A Digital Twin is such a concept that encapsulates many of those technologies.

✉ Xueliang Li
bruce.li@jiangnan.edu.cn

✉ Krist V. Gernaey
kvg@kt.dtu.dk

1 Department of Chemical and Biochemical Engineering, Process and Systems Engineering Center (PROSYS), Technical University of Denmark, Building 228 A, 2800 Kgs. Lyngby, Denmark

2 School of Biotechnology, Jiangnan University, Wuxi 214122, Jiangsu, China

3 Department of Process Engineering, TH Nürnberg, 90489 Nuremberg, Germany

Grieves' concept of Digital Twin was conceived around 2002 for Product Lifecycle Management and later adapted by NASA to describe the development of advanced high-fidelity simulations used for flight certification and flight testing [1]. Several industrial sectors, such as aircraft engine and maintenance suppliers [2] (e.g., GE aviation or Lufthansa Technik), discrete manufacturing suppliers (e.g., Siemens), or the automobile industry, are already benefiting from the increased knowledge achieved after implementing Digital Twins of their assets and systems [3]. For example, Digital Twins are now used in predictive maintenance, where data gathered from key plant equipment and a Digital Model (see below for definition) of that equipment is used to predict time to "failure". This information, in turn, can be used by decision-makers to schedule maintenance as required. This represents a significant improvement over the traditional solution of scheduling maintenance at fixed intervals, without considering the actual state of the equipment. Similarly, in fermentation operations, a Digital Twin can be used to predict batch end times based on data in real-time [4], while also allowing for multiple operational strategies to be tested in silico. Both scenarios lead to increased understanding of how a batch is likely to develop hence allow for informed decision making either in real-time as the process progresses, or "off-line". Clearly, the development of the different facets of the Digital Twin technology promotes a drastic change in the lifecycle of numerous products that extends from the optimisation of the manufacturing process to improved maintenance, risk assessment, and failure detection [3, 5]. Examples of Digital Twin solutions that encompass all these aspects can be found in many industries, including precise manufacturing such as Rolls-Royce's IntelligentEngine program, which allows for improved design, manufacturing, and monitoring of its aircraft engines [6].

Before delving into what constitutes a Digital Twin in the domain of biomanufacturing, a more general idea of the Digital Twin concept can be built up by looking at different aspects and use cases. For example, it is common practice to layout 3D CAD (Computer-Aided Design) models of all types of assets for structural design (during the initial design phase) or later for redesign as part of continuous improvement or a revamping procedure. This enables the virtual inspection, for instance, for assembly and maintenance work. Closely connected is the development of a quick access database of a "digital paper trail" that records and documents all considerations made and standards set during the design, building, and commissioning phase of an asset [7]. In the domain of biomanufacturing the term "Digital Twin" is scarcely used, in contrast to the term "model" (Fig. 1), which will form the backbone of any digital twin as discussed in the following text. Yet if one looks beyond research articles and in particular towards automation and instrumentation vendors that supply solutions to the biochemical industries, relevant content can readily be found in the open literature [8–11]. However, this content mostly follows the conventions adopted by the process control and automation community [3], which is very generic and conceptual.

The definitions used by the control and automation community identify three levels of implementation: (1) a Digital Model, (2) a Digital Shadow, and (3) a Digital Twin [3]. While all three implementation levels require a Digital Model, which is defined as an accurate digital representation of a physical object (in the context of process engineering, a validated process model), the communication between the physical object and the model differentiates these three levels of implementation.

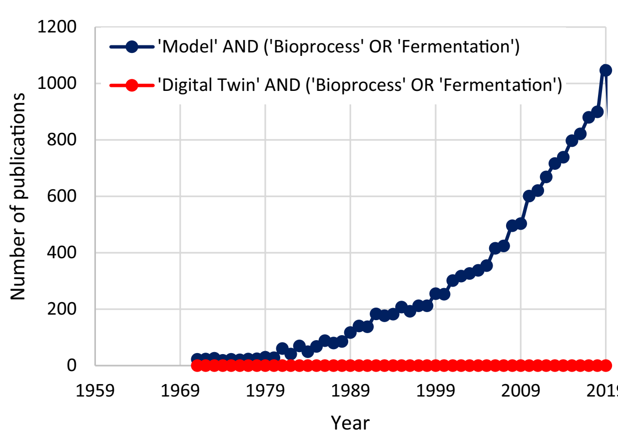

Fig. 1 Number of publications with the terms 'model' and 'fermentation' or 'model' and 'bioprocess' in the title or abstract between January 1971 and May 2020. Source: https://app.dimensions.ai

First, a Digital Model does not communicate in real-time with the physical object. In contrast, data gathered from the physical object can be used for setting up and validating the Digital Model. Second, a Digital Shadow can receive information from the physical object in real-time, but it does not exhibit the ability to communicate back to the physical object. Hence, data flow only in one way. Third, a Digital Twin has the capability to communicate with the physical object bidirectionally. It receives information and communicates back in real-time, thus influences the physical object in real-time. In practice, however, many attempts that would fall into the category of a Digital Model or a Digital Shadow are also referred to as a "Digital Twin" [3–5]. Nonetheless, from a process engineering point of view, this definition of a Digital Twin is significantly skewed towards process control, where the defining feature is the ability of the Digital Twin to act as a closed-loop model-based controller.

Despite the limited literature on Digital Twin in biomanufacturing and the scarce implementation of the concept in practical systems, various elements of the Digital Twin technology can be found in applications related to the real-time monitoring and control of fermentation processes. In this context, the objective of this manuscript is to identify the features a Digital Twin can possess, as well as define and demonstrate its use within fermentation operations. To this end, this manuscript will provide an initial illustration and a description of the elements that constitute a Digital Twin, based on the process automation definitions. This will be followed by a literature review to illustrate the current state of Digital Twin solutions used in fermentation operations and control. Next, a five-step pathway towards the implementation of a Digital Twin (limited to fermentation operations) will be proposed, followed by a lab-scale case study. Finally, the practical implications and benefits of implementing a Digital Twin will be discussed.

Key elements of a Digital Twin

In the context of biomanufacturing operations, a Digital Twin is defined as a comprehensive digital representation of a physical object capable of bidirectional communication with the physical object [3]. As such, the Digital Twin is responsive to changes in the physical asset and can modify the behaviour of the real system [2, 3, 12]. Applying the common definition of Digital Twin established by the process automation community to fermentation operations, Fig. 2 can be developed. This definition closely resembles that of an operator training simulator (OTS) as its capabilities lend well to be a Digital Model.

For an operational Digital Twin to be created, three main components must be developed and then integrated.

- The Physical Plant: the first component is the actual "steel and concrete" plant that a Digital Twin is based upon. This physical twin also contains the sensors and actuators that feed the Digital Twin process data and to act on the process, respectively. To this end, having both information-rich sensors in place that can be used to determine the current state of the unit operations as well as automated actuators on manipulated variables is a pre-requisite for an operational Digital Twin. Today's biomanufacturing plants have partially addressed this pre-requisite. There is an increasing body of work looking into multiple information-rich analytics in addition to the standard temperature, pressure, and flow sensors [13, 14].
- Digitalisation Infrastructure: the second component of an operational Digital Twin is the IT infrastructure needed to connect the digital and the physical part of the plant. This encompasses all the automation aspects in the physical plant and the communication tie-in of the digital plant. Today's biomanufacturing plants allow at least some parts of the production process to be automatically controlled through a SCADA system (Supervisory Control and Data Acquisition) or a DCS (Distributed Control System). Besides, many plants also have a process historian that can be accessed through internal servers. Although newer generations of DCS' are more than capable of facilitating the operation of a Digital Twin, a separate node (dedicated server and information management system) may be needed for its execution. There is also a recent trend in biomanufacturing to migrate data to the cloud [13, 15], although there are still ongoing internal debates about perceived cybersecurity threats.
- Digital Model: the final and the most crucial part of an operational Digital Twin is the Digital Model of the plant, which can take information in real-time and runs concurrently with the real plant. For this purpose, the Digital Model needs to be a validated high-fidelity mathematical model of the biomanufacturing plant that can make predictions, which are then used to adjust the physical plant's operations.

Consequently, the data flow to and from the Digital Model through the digitalisation infrastructure is a key prerequisite in obtaining a Digital Twin. Enabling data flow into the Digital Model (unidirectional data flow) seems to be an achievable goal, which turns it into a Digital Shadow [3, 4]. This allows for dynamic process forecasts to be made in real-time and for making these forecasts available to the operators, e.g., through a suitable visualisation interface. This flow of information allows the operators to act on the process based on the predictions made, i.e., an 'operator-enabled' Digital Twin is achievable. However, the availability of information-rich data in real-time is still somewhat limited in biomanufacturing, and in particular in fermentation operations [4, 15].

When looking at the concept of Digital Twin "at a glance", it can be seen that there is no human intervention required in "closing the loop" between the physical and the digital processes. From a routine operations point of view, this type of fully automated decision making resembles a model predictive controller (MPC). Nonetheless, enabling bidirectional communication (ability to take data and give instructions back to the actuators) between the physical and the digital systems [16, 17] is all but impossible in many industrial fermentation operations as some steps in the process are still carried out by operator intervention. As such, the application of the above definition of a Digital Twin does not consider the characteristics of biomanufacturing, in particular considering that many actions need to be carried out by plant operators due to a lack of automation. Hence, there is a need to re-align the definition of Digital Twin to make it applicable to fermentation operations. A key element in this alignment is the role of the Human Machine Interface (HMI). An effective HMI facilitates the communication between the staff and the Digital Model as both engineers and operators need to interact with it, the one reason being the physical operation interaction mentioned

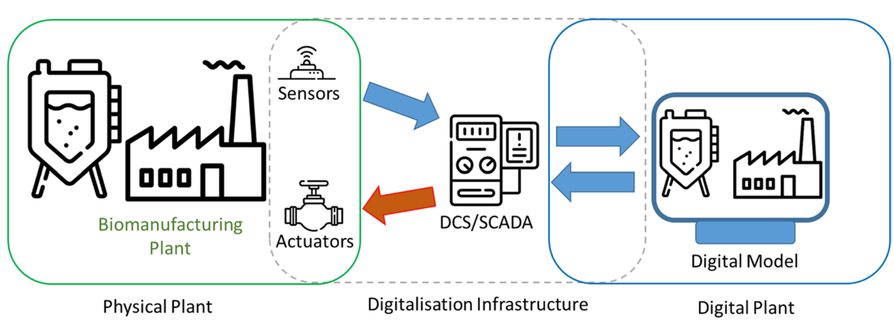

Fig. 2 Illustration of a Digital Twin in the context of biomanufacturing

before, the other being uncertainties in the measurements and the model. As a result, such an HMI should be designed so that the engineers are explicitly made aware of underlying assumptions and estimations of critical operational parameters and their ramifications. The ultimate aim is to ensure that engineers fact-check the outputs against their engineering judgment. From an organisational point of view, the operations team is responsible for ensuring operational safety while meeting product specifications and quotas. Hence, even if a fully automated plant-wide decision-making capability is implemented, plant operators still need to be able to interact with the Digital Model to familiarise themselves with it and to inspect upcoming control moves. Besides, due to the limitations arising from automating functionality in the biomanufacturing industry, it is expected that a fully-fledged Digital Twin will at least partially act through plant operators so that the HMI design will play an even more important role.

Digital Twin technology for biomanufacturing

Accurate mathematical representations of a physical asset (e.g., a unit operation) are central to the development of a Digital Twin. In the biomanufacturing industry, mathematical models are increasingly being used to design and optimize different processes and unit operations [18–21], which could in principle represent the Digital Model element of a Digital Twin. For fermentation operations, a Digital Model begins with the development of a process model that describes the relationship between the feed inputs and the product outputs.

Unidirectional process models in fermentation

Models reactive to process data can be found in numerous implementations of model-based monitoring approaches, where in-line/on-line process data is fed to the model algorithm. Two alternative approaches are generally used to incorporate the real-time data into the model or to the model predictions: (1) state estimation relies on joining a model (which usually has parameters) and measurement data. In this context, the model parameters are fixed, but additional information that becomes available through the measurements is incorporated with the model outputs. (2) Recursive parameter estimation in this context refers to the process state that is determined directly from the model, while data is used to regularly update the model parameters [22].

Recursive parameter estimation (RPE) consists of iteratively updating the parameters of a mechanistic model (e.g., specific growth rates, or inhibition constants) by fitting it to process data collected during the fermentation process. For example, this approach has been validated in the following cases: (1) at a lab-scale lactic acid bacteria fermentation from real-time measurements of pH and ammonia addition [23]; (2) in lignocellulosic ethanol processes using on-line measurements of the glucose concentration measured using spectroscopy and partial least squares (PLS) regression [4]; and, (3) at pilot-scale filamentous fungi fermentation (550 L) using the ammonia addition, the carbon evolution rate (CER) and the oxygen uptake rate (OUR) [24]. Although RPE has been successfully applied to monitor different fermentation processes, the measured data quality can limit its application as RPE does not by default differentiate between measurement noise and the actual process deviations.

State estimators such as Kalman filters are often considered a more robust approach than RPE because they merge process data with the model predictions without altering model parameters. Numerous examples of different state estimating algorithms applied to fermentation processes are described in the literature [14, 25–30]. Successful implementation of state estimation at bench scale (15 L) was developed by Krämer et al. [14, 29] using an extended Kalman filter (EKF) [14] and a sigma point Kalman filter (SPKF) [29]. The authors estimated the states of yeast fermentation by combining on-line and at-line measurements of the optical density, carbon dioxide, pH and spectroscopic measurements of the substrates and product concentrations. Lopez et al. [30] also used an EKF to monitor the concentrations of glucose, xylose and ethanol from ATR-MIR spectroscopy in second-generation ethanol fermentation processes. Golabgir and Herwig, as well as Kager et al. [27, 28], used a particle filter to estimate the process states and kinetic rates in *Penicillium chrysogenum* fed-batch fermentation.

Bidirectional process models

Although the cases outlined above are fundamental for implementing a Digital Twin in fermentation processes, they cannot be considered full Digital Twins because the data flow is unidirectional [17]. Examples of bidirectional data flow between the physical and the digital systems in the biomanufacturing industry are limited and are mainly found in model-based control applications (Table 1).

Mears et al. [24] applied a dynamic model of a fermentation process, updated with real-time measurements (the ammonia addition, the CER and the OUR) and coupled it with a proportional controller to adjust the feed rate in a 550 L fed-batch fermentation. Ehgartner et al. [31] used a capacitance probe and a simple growth model to estimate the biomass concentration and the growth rate of *Penicillium chrysogenum.* They coupled it to a PID controller to adjust the feed rate in fed-batch penicillin fermentation. From a functional point of view, the previous two examples are somewhat similar to the definition of

Table 1 Recent implementations of different levels of a Digital Twin in fermentation systems

System	Volume	Model fidelity	Communication	References
Lactic acid fermentation	2 L	++	Unidirectional	[23]
Lignocellulosic ethanol with yeast	2 L	++	Unidirectional	[4]
Enzyme production with yeast	15 L	++	Unidirectional	[14]
Enzyme production with yeast	15 L	++	Unidirectional	[29]
Penicillin production	15 and 30 L	++	Unidirectional	[27]
Penicillin production	10 L	++	Unidirectional	[28]
Enzyme production	550 L	+	Bidirectional (feed-rate control)	[24]
Penicillin production	10 and 20 L	+	Bidirectional (feed-rate control)	[31]
Penicillin production	2.7 L	++	Bidirectional (feed-rate control)	[32]
Yeast production	Not validated	+++	Bidirectional (feed-rate control)	[33]

The model fidelity is assessed based on the level of detail of the model: (+) corresponds to simple Monod expressions and stoichiometric models, (++) corresponds to empirical models containing a broader description of key metabolites and their interactions (e.g., inhibition effects), (+++) corresponds to dynamic metabolic models describing key metabolic pathways

Note: it needs to be pointed out clearly that none of the work cited in this table has modelled the process equipment and its limitations

a Digital Twin as given by the process automation community. However, these examples do not use high fidelity representations of the inner processes. An example of closed-loop control using a high-fidelity model can be found in [32]. The authors incorporated CER and OUR on-line measurements into a mechanistic model of the fermentation using a particle filter to estimate the biomass formation, the nitrogen, and substrate concentration. These estimations were then coupled to a PID controller and a model predictive controller (MPC) to adjust the feed of different substrates in *Penicillium chrysogenum* fed-batch fermentations. An interesting case of MPC using a high-fidelity model was proposed by Weberhof et al. [33], who combined real-time measurements of volatile compounds in the off-gas with a dynamic metabolic flux balance analysis to describe critical metabolic pathways of *Saccharomyces cerevisiae*. Then, they suggested a strategy to implement this model as an MPC to control the feed rate in fed-batch yeast fermentation. Although, to the best of our knowledge, this approach has not yet been validated experimentally, the detailed description of the fermentation kinetics renders it the closest to a full Digital Twin in bio-based processes, according to the above definition. However, the question may be raised whether a process model, even if it is highly sophisticated, such as the ones based on metabolic pathways, is a high-fidelity model of the asset in the sense of a Digital Twin.

Related concepts

Inferential measurements and soft sensors

Inferential measurements estimate the value of process variables by measuring other parameters that are easier to determine. Soft sensors, also known as hybrid or model-based sensors, are inferential sensing technology built upon the premise that on-line sensors can be used as input signals to algorithms that use mathematical models, computation methods, and prior knowledge to acquire new information [34]. Hardware sensors collect/generate on-line data by monitoring the (bio)processes in real-time; the data is then used in the model to estimate otherwise unmeasured parameters. Those parameters can then be used for monitoring purposes or even be employed in feedback control. In industrial settings, soft sensors must be based on hardware that is robust and straightforward to operate and models that are realistically validated [34, 35]. Ideally, among other benefits, soft sensors should lead to a simpler analytical system, decreased operational and maintenance effort, reduced equipment costs, and bring about a greater degree of automation. In this way, the benefits are well aligned with the PAT initiative. Among other goals, by using PAT technologies, the FDA aims to, for example, increase automation and facilitate continuous manufacturing. Soft sensors can significantly help on both fronts. They can facilitate industrial automation through the use of sensor signals and subsequent derivation of variables that can then be used for feedback control. Hence, they can be seen as crucial enablers for developing Digital Twins. Furthermore, by providing key insights continuously to the following process steps, soft sensors render continuous manufacturing viable [34, 36, 37].

Soft sensors can be broadly classified into two categories, model- or data-driven. The former approach is built upon first-principles models, which entails an in-depth understanding of the (bio)process mechanisms behind it, and is generally considered computationally expensive. In contrast, data-driven models can be quickly developed once they exploit historical data and do not need prior process knowledge. The most commonly used data-driven approaches rely on statistic models (black box models) based on multivariate data analysis (MVDA)—also known as chemometrics. MVDA handles multiple variables simultaneously to extract information from the data by processing and decreasing the dataset's complexity (often the dimensionality). MVDA techniques are used for interpreting analytical data and/or signals from different instruments, such as mass spectrometers, NIR, and MIR, among others. These spectroscopic sensors are especially interesting as hardware for soft sensors since they provide data that are superimposed in continuously updated spectra while these spectral signals can be decomposed into a group of analytical state variables [34]. For example, as a qualitative MVDA approach consistent with the PAT goals, Principle Component Analysis (PCA) has been used for process supervision, to classify raw materials and batches (e..g, the identification of a golden batch). Examples of the use of PCA for bioprocess measurement and monitoring are given in [38]. Quantitative models often use partial least square regression (PLS) to describe correlations among process variables and spectral data; hence it is possible to predict different variables from these measurements on-line [39]. An example of the use of PLS is given in [40].

Artificial neural networks (ANN) form a very popular alternative to MVDA, which is easily set up due to previously established and off-the-shelf methods owing to the technology's popularity in other fields. For example, in [41] the authors apply a dynamic ANN for fermentation monitoring. Moreover, both ANN and MVDA approaches are flexible and can bring several alternatives for predicting significant data in line with the PAT objectives. Other data-driven algorithms used in data-driven soft sensors are, for example, fuzzy logic and support vector regression (SVR) [42]. More recently, the authors in [43] applied deep learning to determine key fermentation model parameters.

Furthermore, model-driven soft sensors based on first principles (mechanistic models, white box models) are often chosen since they depict the bioprocesses more accurately. Although they are easily implemented in the soft sensor, they come with challenges such as determining model parameters from the available analytical data. Examples of mechanistic models used as the soft sensor model for bioprocess monitoring are given, for instance, in [44] and [45]. Furthermore, an increasingly popular approach is the use of hybrid models (grey box models), which take advantage of the best of both mechanistic and statistical models. The challenge, in some cases, might be that the model is difficult to validate. Examples of the application of hybrid models in developing soft sensors are given in [46] and [47].

All in all, it is essential to note that, as highlighted above, the successful implementation of these tools yields Digital Twins' facilitators, not only by providing a comprehensive overview of the process but also by enabling rapid fault detection and process automation [39].

Advanced process control and optimization of bioprocesses

Process control and real-time optimization are vital to the productivity and efficiency of bioprocesses. The biomanufacturing industry has been paying more and more attention to advanced control strategies such as model predictive control (MPC). Compared to conventional control practices (e.g., PID control), advanced process control strategies commonly apply real-time optimization of the productivity and economic gain [48]. In the past few decades, numerous examples have depicted the successful use of advanced process control to different bioprocesses, including fermentation processes [32, 49, 50]. Furthermore, a thorough review of control strategies for open- and closed-loop fed-batch fermentation is presented in [51].

Model predictive control (MPC)

Model predictive control (MPC) has been successfully implemented in the chemical and petrochemical industries; thus, it is only expected that it has significant potential in improving bioprocess productivity. Simply speaking, MPC implementation applies dynamic models to adapt the control strategy in real-time [52]. MPC is mainly used for constrained optimization of unit operations [15].

Model identification in MPC is also of importance where commercial implementations of MPC structures typically rely on step test data to develop input/output relationships (model identification) that are extracted to build an MPC [53]. Since the 1990s, there has been an interest in using Neural Networks to identify complex models (non-linear) for use in MPC [54]. Besides, on-line model identification using data-driven methods has been proposed to counter state- and time-varying processes [55]. In those applications, MPC model identification can be based on first-principles models [56], data-driven [57] and hybrid models [58].

Self-optimizing control (SOC)

An example of using data in closed-loop process control is in improving controller performance through self-optimizing or self-tuning algorithms. Such algorithms adjust critical parameters such as the proportional, integral, and derivative actions of a PID controller to improve its performance according to an established metric [59, 60]. In recent years self-optimising controllers have employed advanced data-driven methodologies such as genetic algorithms [59] as well as artificial neural networks [60]. Similarly, adaptive control, which is in principle similar to self-optimizing control, has also employed neural networks to tune PID controllers to adapt to varying scenarios and requirements [61]. While these examples can be considered a form of data-driven control as they directly change the nature of the closed-loop control action taken, the use of an underlying mechanistic control structure (such as a PID controller) makes them hybrid control approaches.

Real-time optimization

Real-time optimization (RTO), also called on-line optimization, continuously evaluates and changes a process' operating conditions to maximize its productivity (under operational constraints) [62]. RTO is usually applied in highly automated production plants [63] following one of two approaches: the two-layer (cascade) structure or the direct on-line optimizing control method [35, 63]. In the first approach, the upper layer executes the optimization to identify optimal set-points for a set of controlled variables. The set-points are then given to the lower layer (controller layer), where several controllers supervise the estimation of the manipulated variables to keep the controlled variables as close as possible to the defined set-points. The latter approach entails that the manipulated variables are used directly as the decision variables of an optimization problem. The objective function is then calculated over a particular prediction horizon (based upon a nonlinear dynamic process model) [35]. Ochoa et al. [64] have compared both approaches by applying them to a bioethanol production process (plant-wide optimization). Among other examples, Zuo and Wu [65] applied a partial real-time optimization using a hybrid neural network to create the process model, and a genetic algorithm was used for the optimization.

Plant-wide optimization

Plant-wide optimization usually involves a chain of operation units and sometimes even different manufacturing plants operating in different locations. Due to the characteristics of large-scale distributed data gathering (big data), the data can be of several types, have different sampling rates, and high volume, which renders plant-wide modelling and monitoring more challenging. For example, large datasets require significant efforts in terms of data management and storage, information extraction and model interpretation. Thus, even though data-driven modelling and optimization of plant-wide processes have become increasingly popular, the big data features make it a complicated task. There are significant contributions for plant-wide control and optimization, especially in the chemical industry, while only a few focus on bioprocesses. For example, Aydin et al. [66] implemented a plant-wide optimization and control strategy to an industrial diesel hydro-processing process. Jeppsson et al. [67] developed a plant-wide control framework for performance evaluation of wastewater treatment to develop a benchmark simulation model. Finally, in recent work in the field of bioprocessing, Prunescu et al. [68] carried out a model-based plant-wide optimization of a large-scale biorefinery.

Constitutive elements of a Digital Twin in biomanufacturing

The process automation community defines Digital Twin as a comprehensive digital representation of a manufacturing asset that responds to the state of the physical asset and modifies its behaviour [3]. From a practical point of view, the application of this strict definition will discount many digitalisation and mathematical modelling attempts that exhibit some characteristics and functionality of a Digital Twin. However, at the same time, it is also important to differentiate between different attempts made at developing a Digital Twin based on their functionality, the level of mathematical complexity, and the requirement of digitalisation infrastructure. For this purpose, we propose a five-step approach to implement a full-fledged Digital Twin in the context of biomanufacturing, as illustrated in Fig. 3. It is important to note that each step reflects a distinct level towards a Digital Twin of fermentation operations, either through an increase in complexity in terms of mathematical operations or digitalisation infrastructure. A comparison can be drawn to a human embryo, where, in the beginning, the embryo is somewhat undifferentiated and small while progressively developing to the point that an autonomous and self-supporting human being is born. This, of course, requires sustained nourishment from the mother. Similarly, a Digital Twin can start with a basic steady-state model and develop into a full Digital Twin with sufficient and sustained input through its development cycle.

Step 1: steady-state model

A steady-state model represents, at its core, a steady-state mass and energy balance of a process and employs either

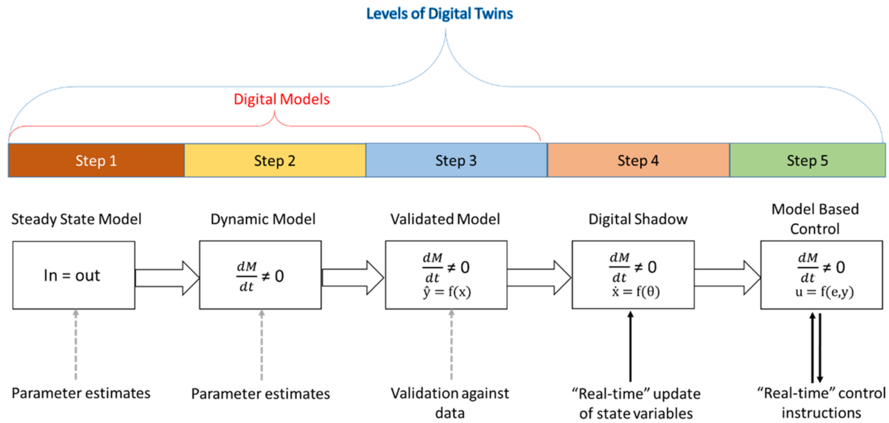

Fig. 3 The steps and actions that must be taken for the transition of a steady-state mathematical model to a fully-fledged Digital Twin

equilibrium conditions to describe reactions or simplified relations to account for the conversion of components. These models are mathematical expressions of a process that are not time-dependent and hence carry no accumulation term. In general, these types of process models are employed in the first pass optimisation and calculation procedures at an initial design stage and for uncertainty analyses [69–71]. In areas such as monoclonal antibodies production, a lack of kinetic information might make steady-state models the only possibility [72].

Step 2: dynamic model

Dynamic models build on the mass and energy balances of a steady-state process model but add accumulation terms and system dynamics. Mathematically, these models contain time-based derivative terms on all variables of interest. In general, these models are employed to identify optimal operational conditions [73–75] and for scaling-up design and process control, e.g., for the identification of operational strategies. It is also important to note that dynamic process models can be purely data-driven, such as the one used for the identification of the end time of a fermentation process [76, 77]. Similarly, hybrid approaches can also be employed where both mechanistic models and statistical models are combined [4, 78, 79].

Step 3: validated model

Validated models extend the capabilities of dynamic process models. The mathematical construct of the model is similar and has time-based derivative terms on all variables of interest. However, there are three additional key requirements to upgrade a dynamic process model to a validated model.

1. The number of states (internal terms the model keeps track of) increases to include terms such as inhibition phenomena, trace substances, side reactions, and interdependencies between states that are not necessarily covered in a dynamic process model. Practically speaking, a validated model must be capable of replicating data sets that have not been used for model identification within a pre-defined band of uncertainty to be considered "validated". At the same time, the model should not output fundamentally incorrect behaviour (e.g. violating a mass balance) within a given operating envelope. By this criterion, a validated model needs to account for a sufficient number of process states to perform accurate predictions on previously unused data sets.
2. A validated model needs to be based on a real "physical process" that is readily identifiable. The model should contain equipment constraint terms such as working capacity and hydraulics. The validated model may also account for valve and sensor dynamics of the "physical process" as they potentially influence the operations of the same.
3. The model needs to be validated against process data obtained from an actual "physical process". This step relies on data-driven techniques ranging from simple regressions [24, 80] to machine learning-based concepts such as differential evolution [81]. In situations where states are directly calculated from measurements, data-driven regression [29, 82] or soft sensors based on engineering fundamentals [83] can be used to estimate these states to aid model validation. Hence, parameters that are used to describe the time-dependent state behaviour, equipment constraints, valve and sensor dynamics must be validated against the physical process.

A few dynamic process modelling endeavours can successfully fulfil the above three criteria to be identified as validated models. Operator training simulators (OTS) are a great example where a dynamic process model is coupled with a graphical user interface (GUI) to train process operators on handling day-to-day or special operational situations as described in several publications [84–88]. Furthermore, Feldman et al. [89] have, for example, employed a validated model of an industrial anaerobic bioreactor for analysis and operational optimization. The development and tuning of a process control structure is another benefit of a validated model [90]. Benchmark simulation models such as the BSM2 (water treatment) [91], the Tennessee Eastman process simulation (petrochemicals) [92], the pharmaceutical simulation (pharmaceuticals) [93] fulfil criteria I and II, and to some extent, criterion III, as these processes are validated against industrial data. A benchmark model can unambiguously be called a validated model if the simulation parameters do not need to be readjusted to account for varying operating conditions, as has been demonstrated for a wastewater treatment plant [94]. A key feature of a validated model is its capability to "predict" with reasonable accuracy how the physical process would behave under a broad range of operations. Data-driven models (including machine learning-based methods) can be a part of a validated model. Even a purely data-driven model is well-suited to be a validated model, as the model development itself covers all relevant phenomena that are identifiable from the data. However, implementing constraints with a data-driven approach may be challenging and exhibits difficulties as constraints may induce strongly non-linear responses, which limits its use.

Step 4: Digital Shadow

A Digital Shadow is a validated model, which can be executed in real-time based on automated input through a data link with the physical process. The Digital Shadow's output is the prediction of the future evolution of the process in real-time. To accomplish the transition of a Digital Model to a Digital Shadow, the following requirements must be fulfilled.

1. Digitalisation infrastructure must be developed, including sensors, data reconciliation, and communication protocols that allow the model to receive data from the physical process in real-time. This includes the development and installation of novel sensor technologies and/or soft sensors that provide information-rich data that can easily be correlated to relevant state information [84, 95, 96]
2. The model must be solved in real-time to ensure timely predictions of the evolving process.
3. A graphical user interface (GUI) is required to communicate the future prediction in real-time to the operators, thus enabling informed actions. Hence, Digital Shadows are primarily an operations support tool. They allow operators to make more informed decisions about a process leading to improved/optimised operations [4, 97]. Figure 4 illustrates the possible information flow from the physical plant to the Digital Model and how a Digital Shadow can aid in operations.

It should also be noted that a Digital Shadow in the context of biomanufacturing will often use advanced monitoring methods that require data-driven models for the parameter and state estimations. This would ensure regular updates of the model to track changes in the physical process. In such cases, a Digital Shadow will fall into the general class of hybrid process models [98, 99].

Step 5: Digital Twin

Finally, a fully-fledged Digital Twin in the context of biomanufacturing, in particular for closed-loop process control and on-line optimisation, can be developed by fulfilling the following criteria:

1. The digitalisation infrastructure, including process automation and communication, allows control structures to be implemented based on the Digital Model predictions. Thus, automated closed-loop control renders operator intervention optional.
2. Implementation of a dynamic optimisation algorithm that can use the Digital Model as its process model and calculate the optimal set of control/operational moves required to optimise overall process economics. For this purpose, a mathematical objective function needs to be determined that captures the overall process economics.
3. The development of digital infrastructure that can guarantee the optimisation (described in criterion II) within

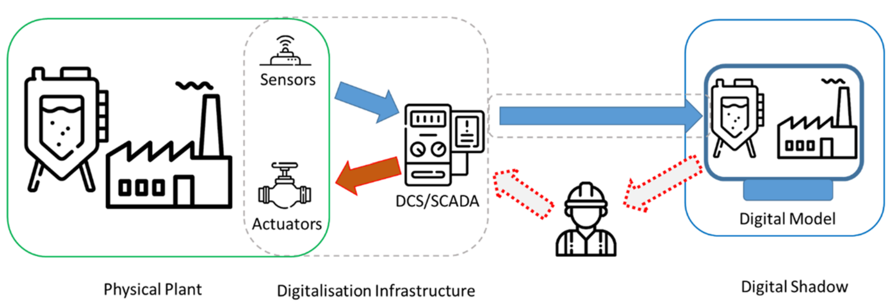

Fig. 4 Representation of a functioning Digital Shadow in biomanufacturing operations

a time constraint; or an alternative arrangement that can guarantee the timely execution, reliability and robustness (e.g., use of a reduced dynamic process model for a given optimisation cycle updated using the Digital Model).

For this specific application (i.e., process optimisation), a Digital Twin has quite a bit of similarity to model-based control (MPC). For example, all criteria listed above also apply to an MPC implementation. However, MPC's typically use transfer functions to describe the relationship between variables, which are simple compared to the high-fidelity system representation of a Digital Twin. Furthermore, the functionality of a Digital Twin is far broader and more detailed than an MPC.

Overall considerations

As discussed at the start of this section, all the five levels of Digital Twin implementation have their distinct functions in the digitalisation of fermentation operations. Conceptually, a Digital Twin implementation (at any level) is not significantly different from established mathematical modelling and process control techniques that are already practised in the chemical and biochemical engineering domains. However, with the increase in computing power and development in digitalisation infrastructure, these Digital Twins, as they evolve during their lifetime, will be more accurate and more robust and offer increased capabilities. Note that the benefit of these increased capabilities has yet to be demonstrated in industrial operations. Nonetheless, the five levels of Digital Twins fit into the overall digital enterprise narrative (of the industry influencers) where steps from conceptual design to process operations are to be fully digitalised.

Human factors in Digital Twins

Reconciling the digital enterprise narrative with biomanufacturing operations, all levels of Digital Twins still need some form of interaction with either the engineers or the plant operators. Hence, the Human Machine Interface (HMI) between the personnel and the Digital Twin needs to be carefully crafted and designed. Engineers working on Level 1 and Level 2 Digital Twins need an HMI that highlights assumptions and estimates made and their ramifications. By contrast, the HMI of a Level 3 validated model should be focused on giving the plant engineers/operators a similar "experience and output" to a typical HMI in a DCS while allowing for efficient scenario analysis. At the Digital Shadow Level 4, the operators need to be made aware of the evolution of key process parameters to take corrective actions, ensuring the process is maintained inside the desired operating envelope. In other words, an HMI needs to support different operational roles.

Case study: second-generation ethanol fermentation

To demonstrate the development of Digital Twins at the different levels listed in the previous section, the lignocellulosic ethanol fermentation described in [4] is used as case-study. Bench-scale batch and fed-batch fermentation of wheat straw hydrolysate were carried out in a 2-L BIOSTAT® A bioreactor (Sartorius, Göttingen, Germany) with pH and temperature control. A genetically modified strain of *Saccharomyces cerevisiae CEN.PK.XXX* was used, which contains a reconstructed metabolic pathway to allow the consumption of glucose and xylose [100].

Step 1: steady-state model

The steady-state model consisted of a mass balance of the different key compounds of the fermentation process. Under anaerobic conditions, *S. cerevisiae* was able to consume glucose and xylose to produce ethanol and biomass (Eqs. 1–2):

$$\text{Glucose} \rightarrow Y_{\text{EtOH/Glucose}} \cdot \text{Ethanol} + Y_{\text{Biomass/Glucose}} \cdot \text{Biomass}, \tag{1}$$

$$\text{Xylose} \rightarrow Y_{\text{EtOH/Xylose}} \cdot \text{Ethanol} + Y_{\text{Biomass/Xylose}} \cdot \text{Biomass}. \tag{2}$$

In addition to the substrates and the products, lignocellulosic feedstocks often contain a high concentration of inhibitors [such as furfural, 5-hydroxymethylfurfural (5-HMF) or acetic acid] derived from the pre-treatment of the biomass. To minimise the toxic effect of the inhibitors, *S. cerevisiae* detoxifies the inhibitors by turning them into less toxic compounds (Eqs. 3–4).

$$\text{Furfural} \rightarrow Y_{\text{Furfuryl alcohol/Furfural}} \cdot \text{Furfuryl alcohol}, \tag{3}$$

$$5-\text{HMF} \rightarrow Y_{\text{Acetic acid/5-HMF}} \cdot \text{Acetic acid}. \tag{4}$$

This steady-state model provides valuable information about the fermentation yields and can be used to estimate the final product titer given a specific substrate composition. However, the steady-state model does not provide any time-dependent information and cannot be used to forecast the fermentation's progress.

Step 2: dynamic model

The authors in [59] used an unstructured microbial kinetic model to describe the dynamic behaviour of lignocellulose to ethanol fermentation processes. The model, containing eight state variables (glucose, xylose, furfural, furfuryl alcohol, 5-HMF, acetic acid, ethanol, and yeast biomass), was defined using Monod-type kinetics and described the growth of *S. cerevisiae* on glucose and xylose, accounting for the inhibitory effects of furfural, 5-HMF, acetic acid and ethanol as well as the effects of catabolite repression (Eqs. 5–11).

Reaction rate

$$v_{\mathrm{Glu}} = -X \cdot \frac{v_{\mathrm{max,Glu}} \cdot \mathrm{Glu}}{K_{\mathrm{S,Glu}} + \mathrm{Glu} + \frac{\mathrm{Glu}^2}{K_{i,\mathrm{Glu}}}} \cdot \left(1 - \left(\frac{\mathrm{EtOH}}{P_{\mathrm{max,Glu}}}\right)^{\gamma_{\mathrm{Glu}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{Fur}}{K_{i,\mathrm{Fur,Glu}}}}\right) \cdot \left(\frac{1}{1 + \frac{FA}{K_{i,\mathrm{FA,Glu}}}}\right) \cdot \left(\frac{1}{1 + \frac{HMF}{K_{i,HMF,\mathrm{Glu}}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{HA}c}{K_{i,\mathrm{HAc,Glu}}}}\right), \tag{5}$$

$$v_{\mathrm{Xyl}} = -X \cdot \frac{v_{\mathrm{max,Xyl}} \cdot \mathrm{Xyl}}{K_{\mathrm{S,Xyl}} + \mathrm{Xyl} + \frac{\mathrm{Xyl}^2}{K_{i,\mathrm{Xyl}}}} \cdot \left(1 - \left(\frac{\mathrm{EtOH}}{P_{\mathrm{max,Xyl}}}\right)^{\gamma_{\mathrm{Xyl}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{Fur}}{K_{i,\mathrm{Fur,Xyl}}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{FA}}{K_{i,\mathrm{FA,Xyl}}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{HMF}}{K_{i,\mathrm{HMF,Xyl}}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{HAc}}{K_{i,\mathrm{HAc,Xyl}}}}\right) \cdot \left(\frac{1}{1 + \frac{\mathrm{Glu}}{K_{i,\mathrm{Glu,Xyl}}}}\right), \tag{6}$$

$$v_{\mathrm{Fur}} = -X \cdot \frac{v_{\mathrm{max,Fur}} \cdot \mathrm{Fur}}{K_{\mathrm{SP,Fur}} + \mathrm{Fur}}, \tag{7}$$

$$v_{\mathrm{HMF}} = -X \cdot \frac{v_{\mathrm{max,HMF}} \cdot \mathrm{HMF}}{K_{\mathrm{SP,HMF}} + \mathrm{HMF}} \cdot \left(\frac{1}{1 + \frac{\mathrm{Fur}}{K_{i,\mathrm{Fur,HMF}}}}\right), \tag{8}$$

$$v_{\mathrm{HAc}} = -X \cdot \frac{v_{\mathrm{max,HAc}} \cdot \mathrm{HAc}}{K_{\mathrm{SP,HAc}} + \mathrm{HAc}} + Y_{\mathrm{HAc/HMF}} \cdot v_{\mathrm{HMF}}, \tag{9}$$

$$v_{\mathrm{FA}} = -Y_{\mathrm{FA/Fur}} \cdot v_{\mathrm{Fur}}, \tag{10}$$

$$v_{\mathrm{EtOH}} = -Y_{\mathrm{EtOH/Glu}} \cdot v_{\mathrm{Glu}} - Y_{\mathrm{EtOH/Xyl}} v_{\mathrm{Xyl}}, \tag{11}$$

$$v_X = -Y_{X/\mathrm{Glu}} \cdot v_{\mathrm{Glu}} - Y_{X/\mathrm{Xyl}} \cdot v_{\mathrm{Xyl}}. \tag{12}$$

The model consists of 8 ordinary differential equations with 32 parameters. This model encompasses all reactions and species of interest and hence serves the purpose of a dynamic model. The parameters found by Mauricio-Iglesias et al. [25] were used as the default parameters, which allowed to develop initial insights into the behaviour of the process and to carry out planning activities for the initial experiments, e.g., concerning batch operating times and default flow rates for the fed-batch fermentation attempts.

Step 3: validated model

While the dynamic model (developed at Step 2) accounted for the main reactions, initial experimental data did not agree with the model prediction [4]. In particular, the actual fermentations were much faster than predicted, indicating that a rate-limiting factor was missing from the model. This was probably due to that the media was supplemented with complex nitrogen sources such as yeast extract and peptone instead of ammonia as used in [25]. Ideally, the nitrogen uptake kinetics should also be included in the model. However, an accurate analysis of the organic nitrogen sources is problematic. To improve the model accuracy, the specific uptake rates of glucose, xylose, and acetic acid were re-estimated (Eqs. 5–6, 9) by fitting the model to off-line measurements collected with high-performance liquid chromatography using the non-linear least-squares method [4]. In this way, the nitrogen source uptake kinetics was accounted for implicitly. Besides, the model was further improved through the incorporation of capacity limitations. Then, the validated model accounted for all relevant process and equipment dynamics to function as a testbed for process monitoring and control activities. This hypothesis can be validated by carrying out a fermentation on a chemically defined medium.

Step 4: Digital Shadow

Transitioning from a validated model to a Digital Shadow involves incorporating real-time data into the kinetic model to account for the evolution of the actual fermentation process. This step is fundamental as Digital Models per se are not able to account for process deviations and might not adequately describe the actual process. However, implementing real-time monitoring tools in lignocellulosic fermentations is challenging due to the high concentration of suspended solids and the media's complex composition [96]. Among the available real-time monitoring methods for fermentation processes, attenuated total reflectance mid-infrared (ATR-MIR) spectroscopy is an attractive option since it allows measuring several key compounds dissolved in the media, limiting the effects of the suspended solid compounds.

Cabaneros et al. [4, 30] used ATR-MIR spectroscopy to measure on-line the concentration of key state variables using a recirculation loop connected to a flow cell. From a practical point of view, the real-time measurement of glucose holds the key to ensure an optimised fermentation operation. The state variables (glucose concentration in [4] and glucose, xylose and ethanol concentrations in [30]) were calculated from the collected spectra using a partial least squares (PLS) regression model. Even though the spectra were turned into actionable information, the monitoring platform alone does not constitute a Digital Shadow. The PLS models allow measuring the system's state, but they neither describe how the different state variables are related nor their time dependency. In essence, a PLS model is part of the sensor component of the physical plant and it does not allow making long-horizon predictions of the fermentation process. Therefore, incorporating the real-time measurements with the predictions made by the Digital Model is the basis for developing a full Digital Shadow. Cabaneros et al. implemented two alternative approaches (a recursive parameter estimation [4] and a state observer [30]) to incorporate measurements into the Digital Model predictions. In [4], Cabaneros et al. sequentially combined the PLS regression model with the Digital Model using a recursive parameter estimation approach. This allowed re-estimating the empirical inhibition constants in real-time based on the progress of the actual fermentation. Figure 5A illustrates the results obtained from the Digital Shadow implementation, where the predictions for Glucose, Xylose, and Ethanol are updated continuously over time as more process data becomes available. In this case, Fig. 5 shows the predictions of the glucose, xylose, and ethanol profile development made for a single ethanol production batch at (A.1) 15 min, (A.2) 5 h and 45 min and (A.3) 24 h and 15 min. This approach resulted in better long-horizon forecasts of the evolution of the fermentation process, as the PLS measurements represent a reliable estimate of the real states of the system. However, the authors showed that when measurements deviated from the real state of the system, these deviations were incorporated into the model parameters and propagated to the output, resulting in distorted predictions. This situation is well illustrated in Fig. 5B.1–2, comparing the root mean squared errors of the prediction in two fermentations, one with accurate measurements (fermentation 1, Fig. 5B.1) and one with faulty measurements due to clogging of the sampling line (fermentation 2, Fig. 5B.2). To overcome this issue, Cabaneros et al. [30] implemented an EKF to estimate the states of the system by merging the real-time measurements of the glucose, xylose and ethanol concentration obtained with the model predictions without changing the model parameters. The authors showed that even though the predictions made with the EKF were also affected by the disturbance in the measurements, the EKF was considerably less sensitive to the faulty measurements as shown by the small increase in the RMSE of the predictions (Fig. 5B.1–2). This led to the conclusion that RPE is an appropriate method for systems where measured data are highly reliable, while state estimators such as the EKF are more robust to noise or deviations in the measurements.

Step 5: Digital Twin

The transition of a Digital Shadow to a Digital Twin requires the predictions obtained from the Digital Shadow to be

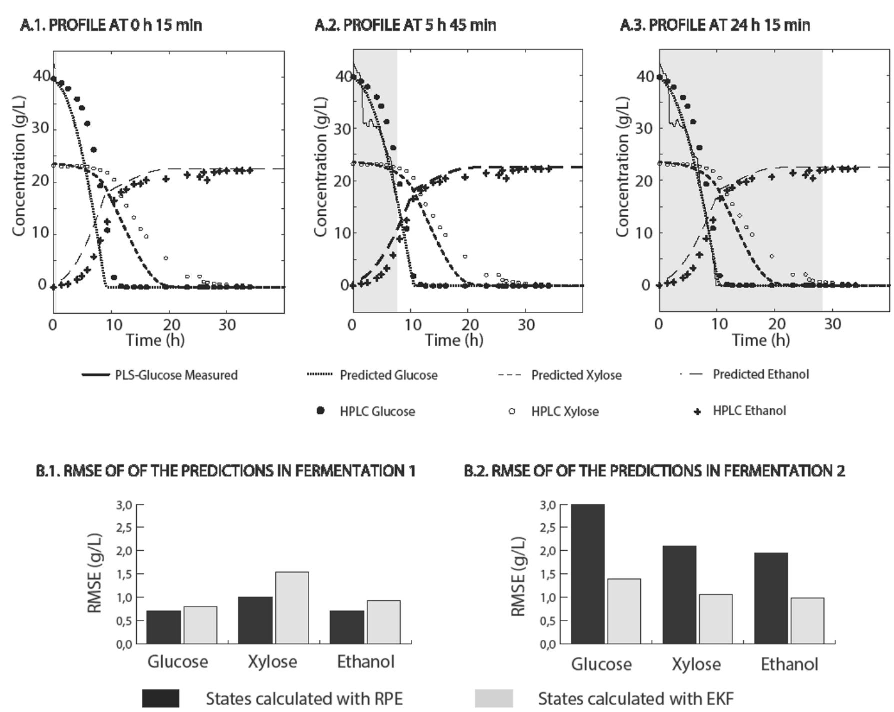

Fig. 5 "Real-time" prediction of glucose, xylose, and ethanol profiles by a Digital Shadow. **A.1–A.3** Model prediction at different time points; **B.1–B.2** model accuracy between two different batches

turned into actions by performing real-time process control calculations to optimise the fermentation process, e.g., by adjusting the feed-rate. The authors have yet to complete this final step in the development of Digital Twin technology. However, in preparation for this, an integrated process monitoring and feed-rate control infrastructure as outlined in Fig. 6 is developed. This infrastructure allows for manipulation of the substrate feed-rate on-line, according to the concentrations of substrates or products measured in real-time using ATR-MIR spectroscopy and PLS regression models. Cabaneros et al. [101] have successfully used this approach to improve the productivity of second-generation ethanol fermentations. A major challenge in such fermentation processes is the strong competitive inhibition that glucose (primary carbon source) exerts on xylose (secondary carbon source), which often results in long fermentations where glucose and xylose are consumed sequentially. To minimize the effects of the competitive inhibition and to promote the co-consumption of the two carbon sources, Cabaneros et al. [101] implemented a closed-loop feedback controller that automatically adjusts the feed-rate of the substrate to keep the concentration of glucose at a given set-point. The controller is based on a PID algorithm and includes a supervisory layer to account for non-linearities in the system [90, 101]. It was demonstrated that operating at low glucose concentrations promoted the co-utilization of the two carbon sources, resulting in a significant increase in productivities (between 20 and 33%, depending on the glucose setpoint).

The development of this control infrastructure is fundamental for the development of a full Digital Twin because it enables bidirectional communication between the physical and the digital systems. However, integrating the Digital Shadow into this infrastructure is still required.

Future perspectives and practical considerations

In the previous section, it was demonstrated how a Digital Twin can be implemented for the application of fermentation process control. This section aims to discuss in brief the practical considerations that must be made when implementing these Digital Twins and how such considerations will shape the adoption of these concepts in the future.

Value proposition

The key driver of digitalisation in an industrial context is the perceived value addition, with or without the ultimate Digital Twin implementation. For example, in the case study, the development of a dynamic model allowed the authors to better understand the expected run time of the fermentation operation and estimates of expected ethanol production and relative concentration. More specifically, the authors demonstrated increased productivities between 20 and 33% as a result of this digitalisation effort. Even though this is not particularly relevant in the context of the case study, having such information in a commercial setting at a detailed design stage facilitates design changes and modifications of the operational doctrine before commissioning. This argument holds for all steady-state as well as dynamic models in fermentation operations. In terms of resources, no infrastructure requirement exists so that these (relatively simple, low-level) Digital Twins can be developed on a reduced budget, using in-house process expertise, providing a sound economic justification for their development.

For validated models, the value proposition is in their ability to operate as an independent testbed that can be used for a multitude of operations. In terms of the case study, the Digital Model was employed to identify optimal operation

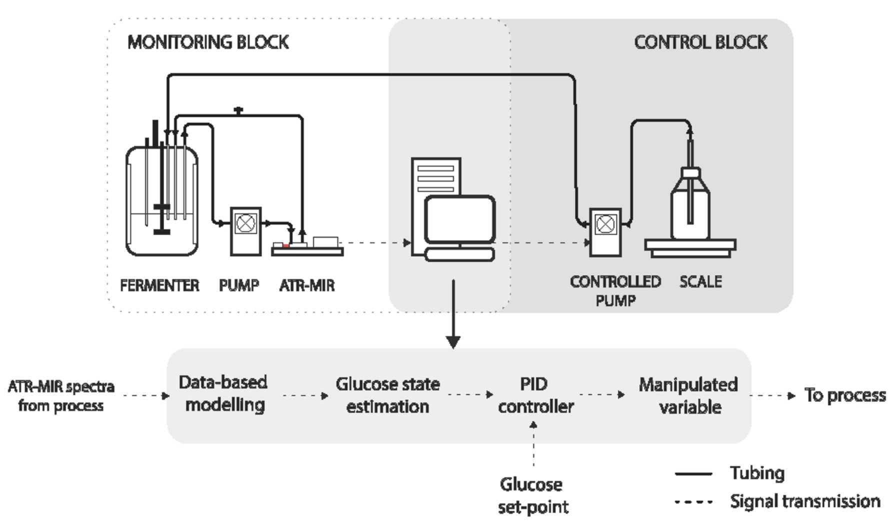

Fig. 6 Schematic representation of the integrated process monitoring and feed-rate control infrastructure developed for real-time process control of lignocellulosic to ethanol fermentation

strategies for given external conditions and to test out advanced process control concepts, including data-intensive control and monitoring methodologies, which are entirely developed on the Digital Model. Although this category of Digital Twin does not have any infrastructure requirement, it does require resources to be allocated for model validation purposes. However, the ability to cut down on the number of physical experiments and the ability to optimise process operations means an economic justification exists for this type of Digital Twin in fermentation operations.

The value proposition for the Digital Shadow and Digital Model-based control is the ability to forecast (and control) the influence of disturbances on the process in real-time that are otherwise difficult to counter. This is particularly relevant for most fermentation operations due to their high degree of variability. However, both concepts require dedicated digitalisation infrastructure to be developed in terms of sensors, process controls, and real-time computing power. Even if a successful implementation provides tangible economic benefits, the noticeable resource allocation requires justification for each case.

Accuracy, robustness and complexity

From a theoretical perspective, one may conclude that the accuracy and complexity of a Digital Twin rise as the level it is categorised in increases, with steady-state models at Level 1 being the least accurate and least complex. In this context, accuracy represents the capability to capture the effects of interest and their evolution in time. If the more complex model accounts for a wider range of secondary effects, it is to be expected that the more complex model proves to be more robust in application terms, contrary to the general belief that simple models are necessarily more robust. Of course, simpler models generally show greater robustness in mathematical terms. Consequently, it is a fine balancing act to perform during the development of a Digital Twin, essentially trading off between accuracy, robustness, and complexity to ensure the developed Digital Twin serves its purpose in a practical application. For example, a relatively simple steady-state model may exhibit greater accuracy and robustness than a dynamic model, particularly when the fermentation in question is not well understood and appears to be chaotic. In this type of situations, it may be highly beneficial to allocate resources primarily to developing accurate and robust steady-state models in contrast to a dynamic model. In operations, the Digital Shadow and a Level 5 Digital Twin may sacrifice both model accuracy and complexity for robustness, as it may be more desirable to be on the correct operating envelope all the time rather than being entirely accurate for most of the time only. Hence, the models used in a Digital Shadow and Digital Model-based control may be less complex than that of a Digital Model developed for the same process. This was apparent in the case study, where a "Digital Model" was dropped in favour of a PLS based state-estimator and a PID based advanced regulatory control loop.

A similar argument can be made about a Digital Twin as well, as ensuring robustness over a larger operating envelope may be more beneficial than model accuracy in a single region. This is particularly valid in operator training, where the model needs to create realistic scenarios to test operators on extreme situations, which requires a relatively high degree of robustness of the model. For controller tuning and design, it may also be desirable to create a balance between model accuracy and robustness. From a practical point of view, a Digital Twin will have to balance such seemingly opposing requirements or add further model complexity to ensure accuracy while maintaining robustness throughout the operating envelope.

Human factor

The human factor plays an important role at all the five levels of Digital Twin, as there are actions (routine or non-routine) that must be taken by humans. In terms of the Digital Models, engineers (mainly from the biochemical disciplines) would have to interact with these models both to solve problems as well as to ensure they are re-validated over time to reflect both operational and structural changes made to the physical plant and unit operations. This is the case for all levels of Digital Twin as some level of maintenance would be necessary to ensure the underlying Digital Model is kept relevant and beneficial. This is in contrast to the general belief where digitalisation is expected to result in job losses in the manufacturing industry [102, 103]. Instead, in the fermentation-based industries, the plant engineer's role would simply change to include responsibilities to maintain these Digital Twins (at all levels) and develop the necessary skill set required to solve complex operational problems.

At the Digital Shadow level, the Human is part of the execution loop where the Digital Shadows' forecasting capabilities must be used by an operator to decide upon and execute corrective actions. Again, the introduction of Digital Shadows will simply change the role of the plant operator where he/she is presented with timely and insightful information that can be used to base decisions on, as opposed to the current industrial practice of following fixed "recipes". However, the fact that an operator can make mistakes (in particular when asked to perform complex tasks under tight constraints) needs to be taken into account by the Digital Model. One option in addressing this shortcoming of humans would be to explicitly model the operators' behaviour in the context of biomanufacturing operations. The development of a "digital operator model" would be a significant undertaking and would require breakthroughs in current thinking. In contrast, concepts such as neural network-based data generation together with further understanding and analysis of operator behaviour may be needed to truly capture operator behaviour in a plant [104]. Besides, the role of the HMI in reducing operator errors needs to be explored.

Conclusions

This manuscript outlines a framework built upon a five-step pathway to define different Digital Twinning attempts relevant in the domain of biomanufacturing. It is based on a combination of Digital Twin concepts from process automation with the current state of the art in fermentation modelling and model-based control. This framework is then applied to a demonstration case that effectively migrates through all five levels of the pathway. Based on both the demonstration case and experiences from the fermentation operation domain, it is determined that understanding (1) the economic value proposition of Digital Twin, (2) the trade-off between model accuracy, robustness, and complexity as well as (3), accurately capturing the effect of plant operators on the manufacturing system hold the key to implementing and further developing Digital Twins. Further, the least complex may not be the most robust model for manufacturing purposes. Finally, digitalisation efforts largely focus on tangible assets and the processes that run inside. But not all manufacturing steps lend well to autonomous operation, so that operators and engineers who interact with the process and/or the model deserve a greater focus in system design, especially concerning the design and implementation of HMIs as they can be converted to a window into the process.

Acknowledgements The Authors would like to acknowledge the Novo Nordisk Foundation in the frame of the 'Accelerated Innovation in Manufacturing Biologics' (AIMBio) project (Grant number NNF19SA0035474).

Funding The work is funded by the Novo Nordisk Foundation in the frame of the 'Accelerated Innovation in Manufacturing Biologics' (AIMBio) project (Grant number NNF19SA0035474).

Availability of data and materials The manuscript does not contain any primary data.

Code availability The work does not contain any custom code nor software application.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest nor competing interests.

References

1. Glaessgen EH, Stargel DS. The Digital Twin paradigm for future NASA and US Air Force Vehicles. In: 53rd Struct. Struct. Dyn. Mater. Conf. Spec. Sess. Digit. Twin, 2011. https://ntrs.nasa.gov/api/citations/20120008178/downloads/20120008178.pdf.
2. Tuegel EJ, Ingraffea AR, Eason TG, Spottswood SM. Reengineering aircraft structural life prediction using a digital twin. Int J Aerosp Eng. 2011;2011:154789. https://doi.org/10.1155/2011/154798.
3. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W. Digital Twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine. 2018;51:1016–22. https://doi.org/10.1016/j.ifacol.2018.08.474.
4. Lopez PC, Udugama IA, Thomsen ST, Roslander C, Junicke H, Mauricio-Iglesias M, Gernaey KV. Towards a digital twin: a hybrid data-driven and mechanistic digital shadow to forecast the evolution of lignocellulosic fermentation. Biofuels Bioprod Biorefin. 2020;14:1046–60. https://doi.org/10.1002/bbb.2108.
5. Lee J, Cameron I, Hassall M. Improving process safety: What roles for digitalization and industry 4.0? Process Saf Environ Prot. 2019. https://doi.org/10.1016/j.psep.2019.10.021.
6. Rolls-Royce. The Rolls-Royce IntelligentEngine—driven by data. https://www.rolls-royce.com/media/press-releases/2018/06-02-2018-rr-intelligentengine-driven-by-data.aspx.
7. Zhang AB, Gourley D. Creating digital collections: a practical guide. Sawston: Chandos Publishing; 2008.
8. ABB. Digital Twin applications. https://new.abb.com/control-systems/features/digital-twin-applications.
9. Siemens. Digital Enterprise as competitive edge. https://new.siemens.com/global/en/company/topic-areas/digital-enterprise/process-industry.html.
10. Honeywell. What is a Digital Twin and why it matters. https://www.honeywell.com/us/en/news/2019/11/what-is-a-digital-twin.
11. Siemens. Empower your data value—with automation digital twin. https://new.siemens.com/global/en/products/automation/industry-software/plant-engineering-software-comos/empower-with-comos/automation-twin.html.
12. Reifsnider K, Majumdar P. Multiphysics stimulated simulation digital twin methods for fleet management. Collect. Tech. Pap.—AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf., p. 1–11, 2013.
13. Gargalo CL, Udugama I, Pontius K, Lopez PC, Nielsen RF, Hasanzadeh A, Mansouri SS, Bayer C, Junicke H, Gernaey KV. Towards smart biomanufacturing: a perspective on recent developments in industrial measurement and monitoring technologies for bio-based production processes. J Ind Microbiol Biotechnol. 2020. https://doi.org/10.1007/s10295-020-02308-1.
14. Krämer D, King R. On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by *S. cerevisiae*. IFAC-PapersOnLine. 2016;49:609–14. https://doi.org/10.1016/j.ifacol.2016.07.235.
15. Udugama IA, Gargalo CL, Yamashita Y, Taube MA, Palazoglu A, Young BR, Gernaey KV, Kulahci M, Bayer C. The role of big data in industrial (bio)chemical process operations. Ind Eng Chem Res. 2020;59:15283–97. https://doi.org/10.1021/acs.iecr.0c01872.
16. Glaessgen EH, Stargel DS. The digital twin paradigm for future NASA and US Air force vehicles. Collect. Tech. Pap.—AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf., p. 1–14, 2012. https://doi.org/10.2514/6.2012-1818.
17. Bevilacqua M, Bottani E, Ciarapica FE, Costantino F, Di Donato L, Ferraro A, Mazzuto G, Monteriù A, Nardini G, Ortenzi M, Paroncini M, Pirozzi M, Prist M, Quatrini E, Tronci M, Vignali

G. Digital Twin reference model development to prevent operators' risk in process plants. Sustainability. 2020;12:1–17. https://doi.org/10.3390/su12031088.
18. Bonvin D, Georgakis C, Pantelides CC, Barolo M, Grover MA, Rodrigues D, Schneider R, Dochain D. Linking models and experiments. Ind Eng Chem Res. 2016;55:6891–903. https://doi.org/10.1021/acs.iecr.5b04801.
19. Narayanan H, Luna MF, von Stosch M, Cruz Bournazou MN, Polotti G, Morbidelli M, Butté A, Sokolov M. Bioprocessing in the digital age: the role of process models. Biotechnol J. 2020;15:1–10. https://doi.org/10.1002/biot.201900172.
20. Mears L, Stocks SM, Albaek MO, Sin G, Gernaey KV. Mechanistic fermentation models for process design, monitoring, and control. Trends Biotechnol. 2017;35:914–24. https://doi.org/10.1016/j.tibtech.2017.07.002.
21. Sinner P, Kager J, Daume S, Herwig C. Model-based analysis and optimisation of a continuous *Corynebacterium glutamicum* bioprocess utilizing lignocellulosic waste. IFAC-PapersOnLine. 2019;52:181–6. https://doi.org/10.1016/j.ifacol.2019.12.255.
22. Dochain D. Bioprocess control. London: Wiley; 2010. https://doi.org/10.1002/9780470611128.
23. Spann R, Roca C, Kold D, Eliasson Lantz A, Gernaey KV, Sin G. A probabilistic model-based soft sensor to monitor lactic acid bacteria fermentations. Biochem Eng J. 2018;135:49–60. https://doi.org/10.1016/j.bej.2018.03.016.
24. Mears L, Stocks SM, Albaek MO, Cassells B, Sin G, Gernaey KV. A novel model-based control strategy for aerobic filamentous fungal fed-batch fermentation processes. Biotechnol Bioeng. 2017. https://doi.org/10.1002/bit.26274.
25. Mauricio-Iglesias M, Gernaey KV, Huusom JK. State estimation in fermentation of lignocellulosic ethanol. Focus on the use of pH measurements. In: Gernaey KV, Huusom JK, Gani R, (eds). Computer aided chemical engineering. Elsevier; 2015. pp 1769–1774.
26. Lisci S, Grosso M, Tronci S. A geometric observer-assisted approach to tailor state estimation in a bioreactor for ethanol production. Processes. 2020;8:480.
27. Kager J, Herwig C, Stelzer IV. State estimation for a penicillin fed-batch process combining particle filtering methods with online and time delayed offline measurements. Chem Eng Sci. 2018;177:234–44. https://doi.org/10.1016/j.ces.2017.11.049.
28. Golabgir A, Herwig C. Combining mechanistic modeling and raman spectroscopy for real-time monitoring of fed-batch penicillin production. Chem-Ing-Tech. 2016;88:764–76. https://doi.org/10.1002/cite.201500101.
29. Krämer D, King R. A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. J Process Control. 2017;82:91–104. https://doi.org/10.1016/j.jprocont.2017.11.008.
30. Cabaneros Lopez P, Udugama IA, Thomsen ST, Roslander C, Junicke H, Iglesias MM, Gernaey KV. Transforming data to information: a parallel hybrid model for real-time state estimation in lignocellulosic ethanol fermentations. Biotechnol Bioeng. 2020;118:1–13. https://doi.org/10.1002/bit.27586.
31. Ehgartner D, Hartmann T, Heinzl S, Frank M, Veiter L, Kager J, Herwig C, Fricke J. Controlling the specific growth rate via biomass trend regulation in filamentous fungi bioprocesses. Chem Eng Sci. 2017;172:32–41. https://doi.org/10.1016/j.ces.2017.06.020.
32. Kager J, Tuveri A, Ulonska S, Kroll P, Herwig C. Experimental verification and comparison of model predictive, PID and model inversion control in a *Penicillium chrysogenum* fed-batch process. Process Biochem. 2020;90:1–11. https://doi.org/10.1016/j.procbio.2019.11.023.
33. Wegerhoff S, Engell S. Control of the production of *Saccharomyces cerevisiae* on the basis of a reduced metabolic model. IFAC-PapersOnLine. 2016;49:201–6. https://doi.org/10.1016/j.ifacol.2016.12.126.
34. Mandenius C-F, Gustavsson R. Mini-review: soft sensors as means for PAT in the manufacture of bio-therapeutics. J Chem Technol Biotechnol. 2015;90:215–27. https://doi.org/10.1002/jctb.4477.
35. Ochoa S, García C, Alcaraz W. Real-time optimization and control for polyhydroxybutyrate fed-batch production at pilot plant scale. J Chem Technol Biotechnol. 2020;95:3221–31. https://doi.org/10.1002/jctb.6500.
36. Rehrl J, Karttunen AP, Nicolaï N, Hörmann T, Horn M, Korhonen O, Nopens I, De Beer T, Khinast JG. Control of three different continuous pharmaceutical manufacturing processes: use of soft sensors. Int J Pharm. 2018;543:60–72. https://doi.org/10.1016/j.ijpharm.2018.03.027.
37. Rato TJ, Reis MS. Building optimal multiresolution soft sensors for continuous processes. Ind Eng Chem Res. 2018;57:9750–65. https://doi.org/10.1021/acs.iecr.7b04623.
38. Glassey J. Multivariate data analysis for advancing the interpretation of bioprocess measurement and monitoring data. In: Mandenius CF, Titchener-Hooker N, editors. Measurement, monitoring, modelling and control of bioprocesses. Advances in biochemical engineering/biotechnology. Berlin, Heidelberg: Springer; 2012. pp. 167–191. https://doi.org/10.1007/10_2012_171.
39. Claßen J, Aupert F, Reardon KF, Solle D, Scheper T. Spectroscopic sensors for in-line bioprocess monitoring in research and pharmaceutical industrial application. Anal Bioanal Chem. 2017;409:651–66. https://doi.org/10.1007/s00216-016-0068-x.
40. Cimander C, Mandenius CF. Online monitoring of a bioprocess based on a multi-analyser system and multivariate statistical process modelling. J Chem Technol Biotechnol. 2002;77:1157–68. https://doi.org/10.1002/jctb.691.
41. Dai X, Wang W, Ding Y, Sun Z. "Assumed inherent sensor" inversion based ANN dynamic soft-sensing method and its application in erythromycin fermentation process. Comput Chem Eng. 2006;30:1203–25. https://doi.org/10.1016/j.compchemeng.2006.02.001.
42. Wang J, Yu T, Jin C. On-line estimation of biomass in fermentation process using support vector machine. Chin J Chem Eng. 2006;14:383–8. https://doi.org/10.1016/S1004-9541(06)60087-6.
43. Gopakumar V, Tiwari S, Rahman I. A deep learning based data driven soft sensor for bioprocesses. Biochem Eng J. 2018;136:28–39. https://doi.org/10.1016/j.bej.2018.04.015.
44. Kiviharju K, Salonen K, Moilanen U, Eerikäinen T. Biomass measurement online: the performance of in situ measurements and software sensors. J Ind Microbiol Biotechnol. 2008;35:657–65. https://doi.org/10.1007/s10295-008-0346-5.
45. Aehle M, Kuprijanov A, Schaepe S, Simutis R, Lübbert A. Simplified off-gas analyses in animal cell cultures for process monitoring and control purposes. Biotechnol Lett. 2011;33:2103–10. https://doi.org/10.1007/s10529-011-0686-5.
46. Teixeira AP, Alves C, Alves PM, Carrondo MJT, Oliveira R. Hybrid elementary flux analysis/nonparametric modeling: application for bioprocess control. BMC Bioinform. 2007;8:30. https://doi.org/10.1186/1471-2105-8-30.
47. Teixeira AP, Oliveira R, Alves PM, Carrondo MJT. Advances in on-line monitoring and control of mammalian cell cultures: supporting the PAT initiative. Biotechnol Adv. 2009;27:726–32. https://doi.org/10.1016/j.biotechadv.2009.05.003.
48. Forbes MG, Patwardhan RS, Hamadah H, Gopaluni RB. Model predictive controlin industry: challenges and opportunities. IFAC-PapersOnLine (Elsevier). 2015;48:531–8. https://doi.org/10.1016/j.ifacol.2015.09.022.

49. Ashoori A, Moshiri B, Khaki-Sedigh A, Bakhtiari MR. Optimal control of a nonlinear fed-batch fermentation process using model predictive approach. J Process Control. 2009;19:1162–73. https://doi.org/10.1016/j.jprocont.2009.03.006.
50. Zhang H, Lennox B. Integrated condition monitoring and control of fed-batch fermentation processes. J Process Control. 2004;14:41–50. https://doi.org/10.1016/S0959-1524(03)00044-1.
51. Mears L, Stocks SM, Sin G, Gernaey KV. A review of control strategies for manipulating the feed rate in fed-batch fermentation processes. J Biotechnol. 2017;245:34–46. https://doi.org/10.1016/j.jbiotec.2017.01.008.
52. Gargalo CL, de las Heras SC, Jones MN, Udugama I, Mansouri SS, Krühne U, Gernaey KV. Towards the development of digital twins for the bio-manufacturing industry. In: Advances in biochemical engineering/biotechnology. Berlin, Heidelberg:Springer; 2020. pp. 1–34. https://doi.org/10.1007/10_2020_142.
53. Udugama IA, Wolfenstetter F, Kirkpatrick R, Yu W, Young BR. A comparison of a novel robust decentralised control strategy and MPC for industrial high purity, high recovery, multicomponent distillation. ISA Trans. 2017;69:222–33. https://doi.org/10.1016/j.isatra.2017.04.008.
54. Wang C, Engell S, Hanisch F. Neural network-based identification and MPC control of SMB chromatography. IFAC Proc. 2002;35:31–6. https://doi.org/10.3182/20020721-6-ES-1901.01156.
55. Akpan VA, Hassapis GD. Nonlinear model identification and adaptive model predictive control using neural networks. ISA Trans. 2011;50:177–94. https://doi.org/10.1016/j.isatra.2010.12.007.
56. Liu C, Gong Z, Shen B, Feng E. Modelling and optimal control for a fed-batch fermentation process. Appl Math Model. 2013;37:695–706. https://doi.org/10.1016/j.apm.2012.02.044.
57. Gomes J, Chopda VR, Rathore AS. Integrating systems analysis and control for implementing process analytical technology in bioprocess development. J Chem Technol Biotechnol. 2015;90:583–9. https://doi.org/10.1002/jctb.4591.
58. Sommeregger W, Sissolak B, Kandra K, von Stosch M, Mayer M, Striedner G. Quality by control: towards model predictive control of mammalian cell culture bioprocesses. Biotechnol J. 2017;12:1600546. https://doi.org/10.1002/biot.201600546.
59. Altınten A, Ketevanlioğlu F, Erdoğan S, Hapoğlu H, Alpbaz M. Self-tuning PID control of jacketed batch polystyrene reactor using genetic algorithm. Chem Eng J. 2008;138:490–7. https://doi.org/10.1016/j.cej.2007.07.029.
60. Kumar V, Gaur P, Mittal AP. ANN based self tuned PID like adaptive controller design for high performance PMSM position control. Expert Syst Appl. 2014;41:7995–8002. https://doi.org/10.1016/j.eswa.2014.06.040.
61. Chang W-D, Hwang R-C, Hsieh J-G. A multivariable on-line adaptive PID controller using auto-tuning neurons. Eng Appl Artif Intell. 2003;16:57–63. https://doi.org/10.1016/S0952-1976(03)00023-X.
62. Naysmith MR, Douglas PL. Review of real time optimization in the chemical process industries. Dev Chem Eng Miner Process. 2008;3:67–87. https://doi.org/10.1002/apj.5500030202.
63. Engell S. Feedback control for optimal process operation. J Process Control. 2007;17:203–19. https://doi.org/10.1016/j.jprocont.2006.10.011.
64. Ochoa S, Repke J-U, Wozny G. Integrating real-time optimization and control for optimal operation: application to the bioethanol process. Biochem Eng J. 2010;53:18–25. https://doi.org/10.1016/j.bej.2009.01.005.
65. Zuo K, Wu WT. Semi-realtime optimization and control of a fed-batch fermentation system. Comput Chem Eng. 2000;24:1105–9. https://doi.org/10.1016/S0098-1354(00)00490-7.
66. Aydin E, Arkun Y, Is G, Mutlu M, Dikbas M. Plant-wide optimization and control of an industrial diesel hydro-processing plant. Comput Chem Eng. 2016;87:234–45. https://doi.org/10.1016/j.compchemeng.2016.01.016.
67. Jeppsson U, Rosen C, Alex J, Copp J, Gernaey KV, Pons MN, Vanrolleghem PA. Towards a benchmark simulation model for plant-wide control strategy performance evaluation of WWTPs. Water Sci Technol. 2006;53:287–95. https://doi.org/10.2166/wst.2006.031.
68. Prunescu RM, Blanke M, Jakobsen JG, Sin G. Model-based plantwide optimization of large scale lignocellulosic bioethanol plants. Biochem Eng J. 2017;124:13–25. https://doi.org/10.1016/j.bej.2017.04.008.
69. Brunet R, Guillén-Gosálbez G, Pérez-Correa JR, Caballero JA, Jiménez L. Hybrid simulation-optimization based approach for the optimal design of single-product biotechnological processes. Comput Chem Eng. 2012;37:125–35. https://doi.org/10.1016/j.compchemeng.2011.07.013.
70. Brunef R, Kumar KS, Guillen-Gosalbeza G, Jimenez L. Integrating process simulation, multi-objective optimization and LCA for the development of sustainable processes: application to biotechnological plants, In: Pistikopoulos EN, Georgiadis MC, Kokossis AC, editors. Computer aided chemical engineering. Elsevier; 2011. pp. 1271–1275. https://doi.org/10.1016/B978-0-444-54298-4.50033-7
71. del Castillo-Romo AÁ, Morales-Rodriguez R, Román-Martínez A. Multi-objective optimization for the biotechnological conversion of lingocellulosic biomass to value-added products. In: Kravanja Z, Bogataj M (eds). Computer aided chemical engineering. Elsevier; 2016. pp. 1515–1520. https://doi.org/10.1016/B978-0-444-63428-3.50257-5
72. Badr S, Sugiyama H. A PSE perspective for the efficient production of monoclonal antibodies: integration of process, cell, and product design aspects. Curr Opin Chem Eng. 2020;27:121–8. https://doi.org/10.1016/j.coche.2020.01.003.
73. Pérez AD, Van der Bruggen B, Fontalvo J. Modeling of a liquid membrane in Taylor flow integrated with lactic acid fermentation. Chem Eng Process Process Intensif. 2019;144:107643. https://doi.org/10.1016/j.cep.2019.107643.
74. Mat Isham NK, Mokhtar N, Fazry S, Lim SJ. The development of an alternative fermentation model system for vinegar production. LWT. 2019;100:322–7. https://doi.org/10.1016/j.lwt.2018.10.065.
75. Da Costa Basto RM, Casals MP, Mudde RF, van der Wielen LAM, M.C. . Cuellar, a mechanistic model for oil recovery in a region of high oil droplet concentration from multiphasic fermentations. Chem Eng Sci X. 2019;3:100033. https://doi.org/10.1016/j.cesx.2019.100033.
76. Spooner M, Kold D, Kulahci M. Harvest time prediction for batch processes. Comput Chem Eng. 2018;117:32–41. https://doi.org/10.1016/j.compchemeng.2018.05.019.
77. Spooner M, Kulahci M. Monitoring batch processes with dynamic time warping and k-nearest neighbours. Chemom Intell Lab Syst. 2018;183:102–12. https://doi.org/10.1016/j.chemolab.2018.10.011.
78. Galvanauskas V, Simutis R, Lübbert A. Hybrid process models for process optimisation, monitoring and control. Bioprocess Biosyst Eng. 2004;26:393–400. https://doi.org/10.1007/s00449-004-0385-x.
79. Cabaneros Lopez P, Feldman H, Mauricio-Iglesias M, Junicke H, Huusom JK, Gernaey KV. Benchmarking real-time monitoring strategies for ethanol production from lignocellulosic biomass.

Biomass Bioenergy. 2019;127:105296. https://doi.org/10.1016/j.biombioe.2019.105296.
80. Gürkan S, Gernaey KV. Data handling and parameter estimation. In: van Loosdrecht MCM, Nielsen PH, Lopez-Vazquez CM, Brdjanovic D, editors. Experimental methods in wastewater treatment. London: IWA Publishing; 2016. pp. 201–234.
81. Wang F-S, Sheu J-W. Multiobjective parameter estimation problems of fermentation processes using a high ethanol tolerance yeast. Chem Eng Sci. 2000;55:3685–95. https://doi.org/10.1016/S0009-2509(00)00038-5.
82. Assis A, Marciel R. Soft sensors development for on-line bioreactor state estimation. Comput Chem Eng. 2000;24:1099–103.
83. Udugama IA, Alvarez Camps M, Taube MA, Thawita C, Anantpinijwatna A, Mansouri SS, Young BR, Yu W. Novel soft sensor for measuring and controlling product recovery in a high-purity, multicomponent, side-draw distillation column. Ind Eng Chem Res. 2019;58:20026–35. https://doi.org/10.1021/acs.iecr.9b04594.
84. Puskás J, Egedy A, Németh S. Development of operator training simulator for isopropyl alcohol producing plant. Educ Chem Eng. 2018;22:35–43. https://doi.org/10.1016/j.ece.2017.11.003.
85. Blesgen A, Hass VC. Operator training simulator for anaerobic digestion processes. IFAC Proc. 2010;43:353–8. https://doi.org/10.3182/20100707-3-BE-2012.0024.
86. Barnert T, Piesik E, Śliwiński M. Real-time simulator of agricultural biogas plant. Comput Electron Agric. 2014;108:1–11. https://doi.org/10.1016/j.compag.2014.06.008.
87. Andersen EB, Udugama IA, Gernaey KV, Bayer C, Kulahci M. Generation for time dependent processes: the tennessee eastman process for generating large quantities of process data. In: 30th Eur. Symp. Comput. Aided Process Eng., 2020.
88. Udugama IA, Gernaey KV, Taube MA, Bayer C. A novel use for an old problem: the Tennessee Eastman challenge process as an activating teaching tool. Educ Chem Eng. 2020;30:20–31. https://doi.org/10.1016/j.ece.2019.09.002.
89. Feldman H, Flores-Alsina X, Kjellberg K, Jeppsson U, Batstone DJ, Gernaey KV. Model-based analysis and optimization of a full-scale industrial high-rate anaerobic bioreactor. Biotechnol Bioeng. 2018;115:2726–39. https://doi.org/10.1002/bit.26807.
90. Udugama IA, Munir MT, Kirkpatrick R, Young BR, Yu W. Side draw control design for a high purity multi-component distillation column. ISA Trans. 2018;76:167–77. https://doi.org/10.1016/j.isatra.2018.03.006.
91. Flores-Alsina X, Solon K, Kazadi Mbamba C, Tait S, Gernaey KV, Jeppsson U, Batstone DJ. Modelling phosphorus (P), sulfur (S) and iron (Fe) interactions for dynamic simulations of anaerobic digestion processes. Water Res. 2016;95:370–82. https://doi.org/10.1016/j.watres.2016.03.012.
92. Downs JJ, Vogel EF. A plant-wide industrial process control problem. Comput Chem Eng. 1993;17:245–55. https://doi.org/10.1016/0098-1354(93)80018-I.
93. Mascia S, Heider PL, Zhang H, Lakerveld R, Benyahia B, Barton PI, Braatz RD, Cooney CL, Evans JMB, Jamison TF, Jensen KF, Myerson AS, Trout BL. End-to-end continuous manufacturing of pharmaceuticals: integrated synthesis, purification, and final dosage formation. Angew Chemie Int Ed. 2013;52:12359–63. https://doi.org/10.1002/anie.201305429.
94. Kazadi Mbamba C, Flores-Alsina X, John Batstone D, Tait S. Validation of a plant-wide phosphorus modelling approach with minerals precipitation in a full-scale WWTP. Water Res. 2016;100:169–83. https://doi.org/10.1016/j.watres.2016.05.003.
95. Pontius K. Monitoring of bioprocesses opportunities and challenges, 2019. PhD Thesis, Technical University of Denmark. https://orbit.dtu.dk/en/publications/monitoring-of-bioprocesses-opportunities-and-challenges-opportuni.
96. Cabaneros P, Feldman H, Mauricio-iglesias M, Junicke H, Kjøbsted J, Gernaey KV. Benchmarking real-time monitoring strategies for ethanol production from lignocellulosic biomass. Biomass Bioenerg. 2019;127:105296. https://doi.org/10.1016/j.biombioe.2019.105296.
97. Montague GA, Glassey J, Ignova M, Paul GC, Kent CA, Thomas CR, Ward AC. Hybrid modelling for on-line penicillin fermentation optimisation. IFAC Proc. 2002;35:395–400. https://doi.org/10.3182/20020721-6-ES-1901.01375.
98. Glassey J, von Stosch M. Hybrid modeling in process industries. 1st ed. Boca Raton: CRC Press; 2018. https://doi.org/10.1201/9781351184373.
99. von Stosch M, Davy S, Francois K, Galvanauskas V, Hamelink J-M, Luebbert A, Mayer M, Oliveira R, O'Kennedy R, Rice P, Glassey J. Hybrid modeling for quality by design and PAT-benefits and challenges of applications in biopharmaceutical industry. Biotechnol J. 2014;9:719–26. https://doi.org/10.1002/biot.201300385.
100. Westman JO, Bonander N, Taherzadeh MJ, Franzén CJ. Improved sugar co-utilisation by encapsulation of a recombinant *Saccharomyces cerevisiae* strain in alginate-chitosan capsules. Biotechnol Biofuels. 2014;7:1–14. https://doi.org/10.1186/1754-6834-7-102.
101. Cabaneros Lopez P, Udugama IA, Thomsen ST, Bayer C, Gernaey KV. Promoting the co-utilisation of glucose and xylose in lignocellulosic ethanol fermentations using a data-driven feed-back controller. Biotechnol Biofuels. 2020. https://doi.org/10.1186/s13068-020-01829-2.
102. McKinsey & Company. An AI Nation?—harnessing the opportunity of artificial intelligence in Denmark. 2019. https://www.mckinsey.com/~/media/McKinsey/Featured%20Insights/Europe/Harnessing%20the%20opportunity%20of%20artificial%20intelligence%20in%20Denmark/An-AI-nation-Harnessing-the-opportunity-of-AI-in-Denmark.pdf.
103. McAfee A, Manyika J. Why every leader should care about digitization and disruptive innovation. n.d. https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/why-every-leader-should-care-about-digitization-and-disruptive-innovation.
104. Lan J, Guo Q, Sun H. Demand side data generating based on conditional generative adversarial networks. Energy Procedia. 2018;152:1188–93. https://doi.org/10.1016/j.egypro.2018.09.157.