Advances in Biochemical Engineering/Biotechnology 177
Series Editor: T. Scheper

Christoph Herwig
Ralf Pörtner
Johannes Möller *Editors*

Digital Twins

Applications to the Design and Optimization of Bioprocesses

Springer

177
Advances in Biochemical Engineering/Biotechnology

Aims and Scope

This book series reviews current trends in modern biotechnology and biochemical engineering. Its aim is to cover all aspects of these interdisciplinary disciplines, where knowledge, methods and expertise are required from chemistry, biochemistry, microbiology, molecular biology, chemical engineering and computer science.

Volumes are organized topically and provide a comprehensive discussion of developments in the field over the past 3–5 years. The series also discusses new discoveries and applications. Special volumes are dedicated to selected topics which focus on new biotechnological products and new processes for their synthesis and purification.

In general, volumes are edited by well-known guest editors. The series editor and publisher will, however, always be pleased to receive suggestions and supplementary information. Manuscripts are accepted in English.

In references, Advances in Biochemical Engineering/Biotechnology is abbreviated as *Adv. Biochem. Engin./Biotechnol.* and cited as a journal.

More information about this series at http://www.springer.com/series/10

Christoph Herwig • Ralf Pörtner • Johannes Möller
Editors

Digital Twins

Applications to the Design and Optimization of Bioprocesses

With contributions by

E. Anane · C. Appl · F. Baganz · T. Becker · M. Canzoneri · M. N. Cruz Bournazou · A. De Luca · D. Eibl · R. Eibl · D. Geier · H. Haße · C. S. S. Hajian · J. Harttung · V. C. Hass · B. Hitzmann · V. Jossen · S. Junne · K. B. Kuchemüller · J. Möller · A. Moser · P. Neubauer · O. Paquet-Durand · R. Pörtner · B. Schumm · R. Sollacher · I. Steinke · R. Takacs · R. Takors · M. Thalhofer · N. Weißenberg · R. Werner · A. Yousefi-Darani · J. Zieringer

Springer

Editors
Christoph Herwig
Institute of Chemical, Environmental and Bioscience Engineering
Vienna University of Technology
Wien, Austria

Ralf Pörtner
Institute of Bioprocess and Biosystems Engineering
Hamburg University of Technology
Hamburg, Germany

Johannes Möller
Institute of Bioprocess and Biosystems Engineering
Hamburg University of Technology
Hamburg, Germany

ISSN 0724-6145 ISSN 1616-8542 (electronic)
Advances in Biochemical Engineering/Biotechnology
ISBN 978-3-030-71655-4 ISBN 978-3-030-71656-1 (eBook)
https://doi.org/10.1007/978-3-030-71656-1

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is divided into two volumes that together provide an overview of the latest advances in the generation and application of digital twins in the field of bioprocess design and optimization. Both tasks have undergone significant transformations over the past few decades, moving from data-driven approaches into the twenty-first century digitalization of the bioprocess industry. Moreover, the high demand for biotechnological products calls for smart and efficient methods during research and development, as well as during tech transfer and routine manufacturing. In this regard, one promising tool is the application of digital twins, which offer a virtual, also known as "in silico", representation of the bioprocess. They mostly reflect the mechanistic of the biological system and the interactions between process parameters, key performance indicators, and product quality attributes in the form of mathematical process models of diverse nature. Furthermore, digital twins allow us to use computer-aided methods to gain an improved process understanding, to test and plan novel bioprocesses, and to efficiently monitor and control them.

In Volume 1 "Digital Twins: Tools and Concepts for Smart Biomanufacturing," a special focus is given to the needs, expectations, and challenges of digital twins in the manufacturing industry. The first chapters focus on the development of digital twins, their economic assessments and the regulatory aspects during industrial implementation. Then, different tools incorporating digital twins are discussed for the design, scale-up, and optimization of bioprocesses.

Volume 2 "Digital Twins: Applications to the Design and Optimization of Bioprocesses" discusses the usage of digital twins in bioprocesses. First, different concepts for digital twin-guided design of experiments are shown, followed by examples for the online implementation of digital twins for bioprocess control strategies. Then, a broad overview about the challenges and opportunities of the implementation of digital twins into existing and newly planned operating value chains are reviewed and their role in the bio(pharma) industry is discussed. In the end, more insights into bioprocesses hydrodynamics and related cellular responses are shown focusing on computer-based methods.

In summary, both volumes provide a diverse and up to date overview about digital twins from well-known scientific and industrial experts. A special focus in each chapter is given to the definition of a "digital twin" due to the individual author's opinion. We believe that both volumes provide a comprehensive compendium of current activities towards the digitalization of bio-manufacturing from research and development up to the final industrial scale. The editors are grateful for the support of all the excellent contributors, the series editor Prof. Thomas Scheper, Institute of Technical Chemistry, University of Hannover (Germany), and the publishers who have made both volumes possible.

Wien, Austria
Hamburg, Germany

Christoph Herwig
Ralf Pörtner
Johannes Möller

Contents

Adv Biochem Eng Biotechnol (2021) 177: 1–28
https://doi.org/10.1007/10_2020_154

Published online: 31 December 2020

Potential of Integrating Model-Based Design of Experiments Approaches and Process Analytical Technologies for Bioprocess Scale-Down

Peter Neubauer, Emmanuel Anane, Stefan Junne, and Mariano Nicolas Cruz Bournazou

Contents

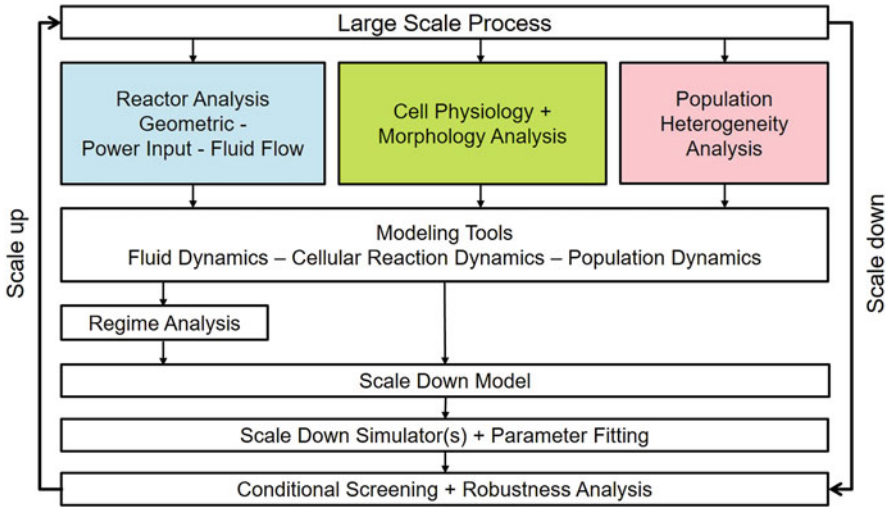
Abstract Typically, bioprocesses on an industrial scale are dynamic systems with a certain degree of variability, system inhomogeneities, and even population heterogeneities. Therefore, the scaling of such processes from laboratory to industrial scale and vice versa is not a trivial task. Traditional scale-down methodologies consider several technical parameters, so that systems on the laboratory scale tend to qualitatively reflect large-scale effects, but not the dynamic situation in an industrial bioreactor over the entire process, from the perspective of a cell. Supported by the enormous increase in computing power, the latest scientific focus is on the

P. Neubauer (✉), E. Anane, and S. Junne
Bioprocess Engineering, Faculty III Process Sciences, Institute of Biotechnology, Technische Universität Berlin (TU Berlin), Berlin, Germany
e-mail: peter.neubauer@tu-berlin.de

M. N. Cruz Bournazou
DataHow AG, Zurich, Switzerland

application of dynamic models, in combination with computational fluid dynamics to quantitatively describe cell behavior. These models allow the description of possible cellular lifelines which in turn can be used to derive a regime analysis for scale-down experiments. However, the approaches described so far, which were for a very few process examples, are very labor- and time-intensive and cannot be validated easily. In parallel, alternatives have been developed based on the description of the industrial process with hybrid process models, which describe a process mechanistically as far as possible in order to determine the essential process parameters with their respective variances. On-line analytical methods allow the characterization of population heterogeneity directly in the process. This detailed information from the industrial process can be used in laboratory screening systems to select relevant conditions in which the cell and process related parameters reflect the situation in the industrial scale. In our opinion, these technologies, which are available in research for modeling biological systems, in combination with process analytical techniques are so far developed that they can be implemented in industrial routines for faster development of new processes and optimization of existing ones.

Graphical Abstract



Keywords Bioprocess scale-up, Process analytical techniques, Process modeling, Scale-down

1 Status of Bioprocess Scale-Down: The Need for a Model-Based Design

Studies on the scale-down of bioprocesses have received much attention in recent years. On the one side, this is due to an increasing implementation of the concept of a circular bio-economy. Within this framework there is a boost in new processes and

strategies. In the field of biotechnological production of basic chemicals and biocatalysts, the reactor volumes are getting steadily larger to ensure the necessary yields. Newly conceived processes must be robustly feasible on a large scale [1]. On the other side, in the field of biopharmaceutical production, product yields in the reactor originally had a lower priority for originator products compared to the costs incurred in downstream processing. The expiry of many patents and the development of generic or biosimilar products has led to a price pressure that companies can only counteract with very efficient and less variable bioprocesses. Additionally, the implementation of single-use strategies, especially in connection with the lower power input in single-use bioreactors, introduces scaling phenomena, i.e. imperfect mixing issues already at much smaller scales, i.e. in reactors with 1–5 m^3 [2]. Still, the main reason for the increased interest in scale-down investigations are failures, lower yields, higher batch-to-batch variability, or even changes in the product quality of processes performed in industrial scale. The difficulties to predict the outcomes at this scale, the large number of unexpected responses of cells to different environments, and the impossibility of computer based tools to foresee the changes on the phenotype throughout scales based on laboratory data are the driving forces behind scale-down popularity [1].

The discrepancy between laboratory- and industrial-scale fermentation processes results from heterogeneous environments in large-scale bioreactors due to the limited volumetric power input and geometric issues, which result in longer mixing times associated with the increasing volumes at industrial scale, compared to laboratory scale bioreactors [3]. The effects of such process inhomogeneity on microbial physiology and product syntheses, including the quantity and quality of recombinant products, have attracted much attention in the bioprocess research community, due to the mostly unforeseeable impacts of the heterogeneities on process efficiency, e.g. by the accidental incorporation of non-canonical amino acids into the product [4].

In the past three decades, various forms of single-compartment and multi-compartment scale-down bioreactors have been developed to study scale-up effects in fermentation processes [3]. This development, however, is accompanied by a constant discussion about the extent, to which these systems really reflect the conditions at an industrial scale. Concrete proposals for procedures for scaling down a process to laboratory scale have only been developed in recent years, see, e.g. [5], but are generally very sophisticated and therefore unsuitable for broad application.

In parallel, during the last years, there has been a phenomenal increase in the use of high-throughput (HT) miniaturized bioreactor systems for strain screening and bioprocess development, which has significantly reduced the times required for early bioprocess development. These new powerful laboratory tools require, however, new methods for planning, performing, and evaluating these highly parallel experiments. The systems are no longer treatable by manual methods – therefore, standard methods of design, mathematics and statistics, modeling and process engineering as they have been used in other disciplines for a long time have to be implemented and adapted in the field of bioprocessing. Intuitively, when dealing with large data sets,

highly automated systems and closely interconnected devices, concepts like Internet of Things (IoT) and Digital Twin come to mind. Beyond the hype around Digital Twins, its history and modern definition are closely related to High-Throughput Bioprocess Development. The term Digital Twin emerged in the field of Product Lifecycle Management to increase the efficiency in product and process development [6]. Mathematical models, used for process monitoring (observers) in feedback loops, for approximate optimal control applications (MPC), or even in real-time optimization have existed for quite some time now [7]. Yet, the extension of these methods including IoT, big data, and fully autonomous systems might require a new terminology [8].

The Digital Twin, envisioned as a mirror image (an exact copy) of the physical system that follows the complete lifecycle of the product from idea to manufacturing, is possible only if (1) an exact representation of the system in mathematical equations is at hand and (2) the current state of all relevant elements of the real system can be fully monitored through real-time data. In bioprocess development, building a Digital Twin implies joining High Throughput, Omics, PAT, Machine Learning, Bioprocess Automation, and Bioprocess Systems Engineering tools to enable the development and operation of a biomanufacturing plant with a perfect copy of all units from the molecular/intracellular level up to large-scale dynamics. Such a Digital Twin is clearly far beyond the capabilities of current technologies. Still, it defines a clear roadmap that shows the relevance of the integration of different fields and tools to maximize the efficiency of bioprocess development. Mathematical models, which form the basis of digital twins, support all fields of biotechnology and bioprocess engineering [9]. This includes biochemical systems [10], systems biology [11], metabolic engineering [12], flux balance analysis [13], synthetic biology [14], and bioinformatics [15]. A good overview of applications of mathematical models, as well as a proof of their slow advance in bioprocess engineering is given by Jay Bailey [16]. Nevertheless, the complexity of biological systems poses difficult challenges to the direct use of advanced mathematical techniques in bioprocess development [17].

The complexity of the underlying metabolic and physiological phenomena demands large nonlinear equation systems with a large number of unknown and often time-variant parameters. The existing methods are too complex and computationally expensive for application in biotechnology [18–20]. Compared to general applications in engineering [21–24], biotechnological applications typically lack sufficient data, as well as process understanding [25–27].

Finally, the advances in artificial intelligence, especially in data-driven learning tools, offer incredible possibilities, but need to be adapted to the specific needs of bioprocesses, which have peculiarities (e.g., evolution of the biological system [28]), broad population distributions, very complex chemical composition and complicated (metabolic) reaction networks that are not present in mechanical or chemical processes [12]. Nevertheless, such mathematical tools have greatly contributed to our understanding of the interactions between the organism and the constraints of growth in bioreactors, as well as the elucidation of otherwise obscure intracellular processes [13].

The problem is, however, that these tools and concepts, namely scale-down bioreactors, high-throughput mini bioreactors, and model-based tools, have mainly developed in parallel, with little or no interaction among them. In fact, scale-down bioreactors are still operated as standalone, low-throughput devices [3]; and the benefits of mathematical models are not fully exploited in both scale-down and high-throughput systems [29]. Therefore, the actual challenge is to combine these very special techniques such that they can work together efficiently.

In this work, we discuss the current state of process development focusing on scale-down, the typically underestimated milestone. We discuss existing experimental tools, sensor technologies, and latest advances in computational methods for the design of scale-down investigations. Next, we demonstrate the issues related to the current decoupled efforts to address process development. Finally, we describe the required steps to reach a proper integration of all tools to create a digital twin of the bioprocess development procedure together with its potential and future applications.

2 The Digital Twin in Bioprocess Development

The answer to the current challenges in advanced bioprocess development (see Fig. 1) is a digital twin that covers all developmental stages and allows an efficient and effortless transfer of knowledge and information throughout the complete process [30]. Thus, the term "digital twin" covers more than just the mathematical model of a single component of the process. With regard to industrial bioprocesses, "digital twins" can describe the biological system itself or parts of the system, e.g. the three-dimensional structure of the protein product. They can also describe phases of process development, such as strain screening, different scales of the fermentation process and downstream operations, and in final production they can be used for the design or installation of a production plan, as well as for the control of the actual manufacturing process including its optimization.

The required advances in automation, process analytical technologies (PAT), and computer-aided tools for bioprocess monitoring and control are available [31]. The main challenge in building a functional digital twin is the difficulty in harmonizing these existing technologies through standardized communication protocols and data management systems. Such a digital framework tightly embedded into the highly automated experimental systems and production facilities through PAT and advanced mathematical modeling tools can build the path for knowledge transfer between the whole bioprocess development workflow.

Scale-up and scale-down present arguably the most descriptive examples for the challenges of knowledge transfer as well as its relevance in bioprocess development [32, 33]. Scaling is basically an effort to transfer the information generated in one stage to another aiming to maximize the generation of relevant knowledge for the industrial process [17].

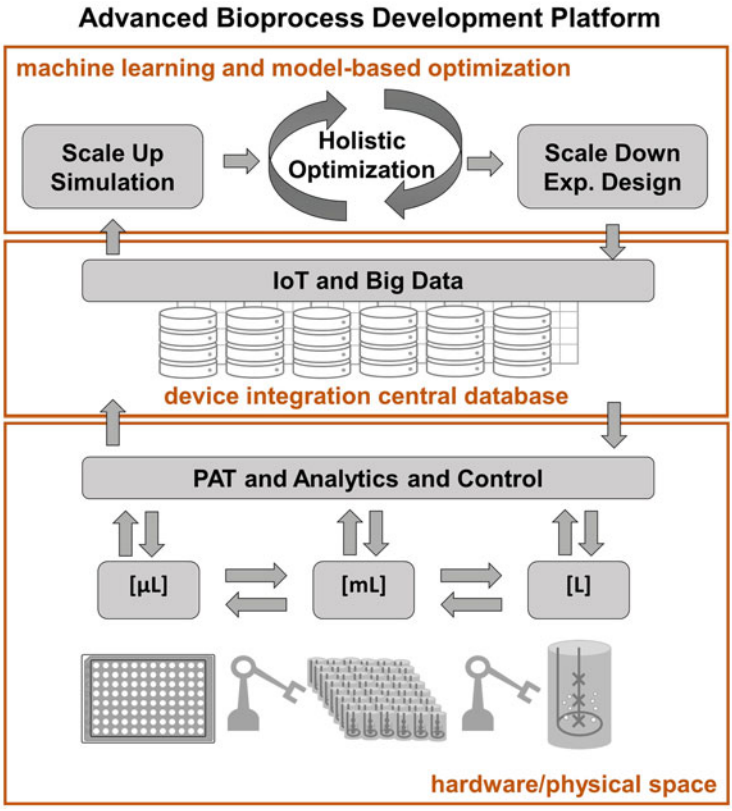


Fig. 1 The role of the digital twin in advanced bioprocess development. Integration of scales, units, and disciplines. From bottom to top, automated hardware (the physical system), device integration, data transfer and handling, and model-based optimization tools for scale-up and scale-down

In biotechnology, scale-down is eminent to assure the generation of relevant-to-process knowledge in lab scale. As discussed further below, the variation to the cellular response caused by the bioreactor stresses cannot be predicted without extensive experimental information under proper conditions. That is, the experiments at lab scale must create the proper environment, to emulate industrial conditions which are unfortunately difficult to predict beforehand due to the highly complex interaction between the organism(s) and the bioreactor. Scale-up is the art of extracting knowledge from experimental data to translate it into an efficient manufacturing strategy. This process is as challenging as scale-down for the same reasons: insufficient data on the underlying dynamics of the bioprocess and the lack of a proper mathematical translation of the information throughout scales. In both cases (scale-up and scale-down), purely data-driven methods fail to understand the complex interactions between the organism(s) and the distinct bioreactor environment, generating unrealistic predictions for the next scales and corroborating the risks of extrapolating black-box models [34, 35]. On the other side, highly complex

mechanistic models offer important insights into the system [36, 37], but require process data that is very expensive, if at all obtainable.

There is no doubt that the high complexity of living organisms, especially in connection with their flexible information networks, which have been developed in millions of years of evolution, is the main challenge in modeling biological processes, and hence the reason why purely in silico development methodologies are doomed to fail [17]. Natural organisms have been so successful in surviving unfavorable conditions as they are characterized by genetic and metabolic heterogeneity between the individuals in a population. This increases especially under stress conditions in an effort to maximize the chance for some to survive [38]. Bioprocesses on an industrial scale (1) contain highly heterogeneous cell populations and (2) induce a constantly changing environment that sets the organisms under a high stress due to the inhomogeneities in the reactor. It is for this reason that advanced experimental facilities, sensor technologies, and mathematical modeling must be tightly integrated into a digital twin framework to finally achieve development times comparable to other industries [39–43]. The predictions at the digital level require a continuous re-calibration and evolving mathematical description to cope with the unpredictable behavior of living systems as well as efficient strategies to design and operate the experimental campaigns [29]. Most importantly, these models also reflect the dynamic development of the heterogeneity of the population and take statistical uncertainties into account in order to account for possible batch-to-batch variations on an industrial scale and be robust in terms of their prediction.

3 Inhomogeneities in Industrial-Scale Bioreactors and Their Influence on the Biological System

Bioprocess development is usually started in shaken cultures, traditionally in shake flasks, or more recently in parallel microwell plates where only a few endpoint measurements are possible [44]. Nowadays there is a great interest in implementing the final strategy, the fed-batch method, very early in the development process. For this purpose, various methods have been developed in the past years, which make this possible despite the small volumes [45].

Although this represents a significant advance on the traditional approach, the methodologies can vary greatly depending on the applied system. Substrate feeding can be either continuous or intermittent (pulse-based), and the controllers for the continuously measured parameters (e.g., pH value and dissolved oxygen) can be set differently, if these parameters are adjusted at all. If, as discussed above, we assume that the culture is highly sensitive to the process conditions – and thus the product formation is influenced accordingly – it is necessary to set these parameters so that the conditions are similar to those on an industrial scale.

This is important, especially for the development of biopharmaceuticals, where drug substance for clinical trials may be produced as early as the laboratory development phase, or in early pilot plant phase under cGMP guidelines. Once regulatory clearance is obtained, the process is then scaled up to industrial conditions. Two important points must be considered here: (1) the product quality characteristics affecting the efficacy of the biopharmaceutical candidate must comply with the specifications defined during clinical trials and (2) the process must be economically feasible whilst producing sufficient quantities to supply the market. Unfortunately, the increase in size has many implications for the process conditions inside the bioreactor. That is, if an 80 L bioreactor is scaled up to 10,000 L whilst maintaining a constant power input per unit volume, the mixing time increases 3 times, the impeller tip speed doubles, and the shear forces increase almost 10 times [46]. Oldshue showed that a scale-up design to satisfy mass transfer (constant $K_l a$ criterion) from a 75 L pilot scale process to a 95,000 L production scale would increase the shear rate by 180%, whereas maintaining a constant shear rate between the two scales could only produce 40% of the mass transfer requirements of the culture in the large scale [47]. The most common consequence is an inevitable increase in mixing times of up to 200 s in larger-scale bioreactors (since scale-up is mostly based on $K_l a$, P/V, impeller tip speed) [48].

In addition to the increased mixing times, fed-batch processes are fed with concentrated substrates at localized feeding points, which are mechanically fixed. The longer mixing times and the localized addition of highly concentrated viscous substrates lead to the formation of concentration gradients in the bioreactor [49, 50]. Cells that are traversing these gradients respond in many ways, by the varied distribution of metabolic fluxes due to the changed uptake rates in different positions of the bioreactor, and by specific gene expression profiles which include both specific responses and general stress adaptation. The specific reaction of an individual cell depends not only on its metabolic state and the current phase in the cell cycle, but also on its specific historical situation, i.e. what conditions it has experienced previously in the dynamic course of time [51]. This is currently being investigated using fluid dynamic models by simulating cell lifelines. The sum of all of these affects the fermentation efficiency in terms of yields and overall process robustness.

When the characteristic time of relevant cellular processes (translation, cell division) is close to the mixing time in large-scale bioreactors, there is a measurable influence of gradients on the growth and metabolic behavior of the culture [46, 52]. The inefficient mixing in large-scale bioreactors leads to the creation of spatial concentration pockets of relevant process parameters, such as substrate (glucose), dissolved oxygen, acidity, and temperature. Furthermore, GMP manufacturing processes suffer from the rigidity of the process due to the difficulty in using validated equipment for such studies, especially when the characterization study requires minor retrofitting of the bioreactor, such as installing extra sensors. In cases where bioreactor characterization has been done, companies consider the data as confidential; therefore, the information is not available to the scientific research community.

Substrate Gradients In fed-batch cultures, the existence of excess substrate zones in the broth defeats the purpose of this tight control for the fraction of the culture that comes into contact with these zones. The exposure of the culture to zones of higher substrate concentrations has direct consequences on the uptake capacities of the cells for this substrate [51, 53]. As a result, the excess substrate zones may cause the cells to grow at the maximum specific growth rate, which may plunge organisms such as *E. coli* and *Saccharomyces cerevisiae* into overflow metabolic states as reported in numerous studies [51, 54–56]. The high metabolic flux of glucose through the glycolysis which is favored by high affinity uptake systems, i.e. low K_S values, also leads to the accumulation of NADH-H^+, and thus to a higher rate of respiration.

As a consequence, the high metabolic activity in the feeding zone can also lead to oxygen limitation if the biochemical reduction of oxygen by the cells is faster than the limited diffusion of oxygen into the cultivation medium. It is likely that the uneven distribution of the substrate due to feeding is the main cause for the *dissolved oxygen gradients*, besides the uneven fluid-dynamic distribution of the gas bubbles. The dissolved oxygen problem which is basically caused by the inherently low solubility of oxygen in fermentation broths [49] becomes even greater in processes with pellet forming organisms (oxygen gradient in the pellet) or shear-sensitive cells (limited sparging to prevent shear stress caused by the bursting of gas bubbles) [57].

Temperature Gradients Temperature gradients are among the least studied scale-up effects in bioprocess development. Although it is clear from a microbiological point of view that small temperature fluctuations of a few degrees have a major impact on cellular reactions and that, from a process engineering perspective, precise temperature control in industrial bioreactors is a serious problem, to the best of our knowledge, there is no information about local temperature profiles in industrial bioreactors, nor have experiments been performed in scale-down simulators to simulate the effect of perturbing temperatures on a process.

pH Gradients pH gradients are recently gaining attention in the bioprocess research community. Simen et al. investigated the effect of ammonia pulses (shifts in pH) in *E. coli* cultivations and observed a higher maintenance energy and the activation of over 400 genes in response to the pH gradients [58]. pH gradients are also relevant in industrial-scale batch cultivations of lactic acid bacteria. This has been revealed by combined approach by the use of multiple pH probes and a computational fluid dynamic model coupled with a kinetic model for a process of *Streptococcus thermophilus* in a 700 L pilot scale bioreactor [59]. Recently, we also could demonstrate by two- and three-compartment bioreactor cultivations that such pH oscillations affect the cocci chain length and decrease the growth rate in *S. thermophilus* cultures (manuscript in preparation). Also in CHO fed-batch bioprocesses pH perturbations decrease the cell viability and increase lactate accumulation [60]. Also pH oscillations have been recently demonstrated to affect product accumulation in a cell line specific manner [61].

Carbon Dioxide Gradients In microbial cultivations, a recent study of CO_2/HCO_3^- gradients in *Corynebacterium glutamicum* showed no significant impact

of these stresses in the physiological response of the organism, although there was a marked increase in the expression of certain genes, upon genomic analysis [62]. In a recent report, *E. coli* cells exposed to CO_2 levels above 70 mbar CO_2 partial pressure in the inlet gas led to reduced biomass yields and rapid accumulation of acetate, even under non-overflow and fully aerobic conditions [63].

Interaction Between Multiple Gradients Finally, the results of Limberg and colleagues show that when pH gradients are coupled to oxygen limitation, *C. glutamicum* loses its robustness against dissolved oxygen fluctuations [64], leading to yield losses of up to 40%. This implies that the study of concentration gradients in fermentation should be conducted in a multi-faceted manner, to consider all possible gradients and the necessary combinations among them to arrive at a more holistic conclusion for each strain. There is also a close correlation between pCO_2 levels, pH, base addition, and osmolality in large-scale CHO cell cultures which affect the metabolic lactate shift (transition from lactate production to lactate consumption) [65, 66].

4 Framework for Bioprocess Scale-Down Studies

4.1 Characterization of the Large Scale

A good characterization of the large-scale bioprocess is important to conclude proper scale-down experiments which really imitate the large scale (see Fig. 2). Since the scale-down data is only as good as the environment it mimicked, it is absolutely necessary to characterize both the cellular state and specific heterogeneity (gradient profiles) in the larger scale. Standard analytical methods of the medium and gas composition and the derivation of cell specific rates need to be complemented by direct monitoring of the physiological state of the cells. A proper scale-down methodology should be based on the similarity of cellular responses, all at the level of metabolism, protein expression, and population heterogeneity between laboratory and industrial scale. In order to avoid false conclusions and to reduce the risk of scale-up, robustness analyses must be used to assess the final batch-to-batch variability. This complex problem can only be solved if digital approaches (digital twin) can be coupled with a large number of experiments.

In the past, there were a large number of approaches to simulate these gradients occurring in the industrial bioprocess in scale-down systems, see reviews by Neubauer and Junne [3], Lara et al. [46], Delvigne and Noorman [67]. All these systems achieve oscillating conditions regarding the specifically investigated parameters, i.e. the specific parameters which were the focus of the investigators. Different priorities were set depending on the specific approach. In multi-compartment reactors, the dominant parameter is the residence time distribution in different compartments where cells are located within a defined period of time. In more-compartment stirred tank systems, the zones are characterized by a previously defined state, e.g. in

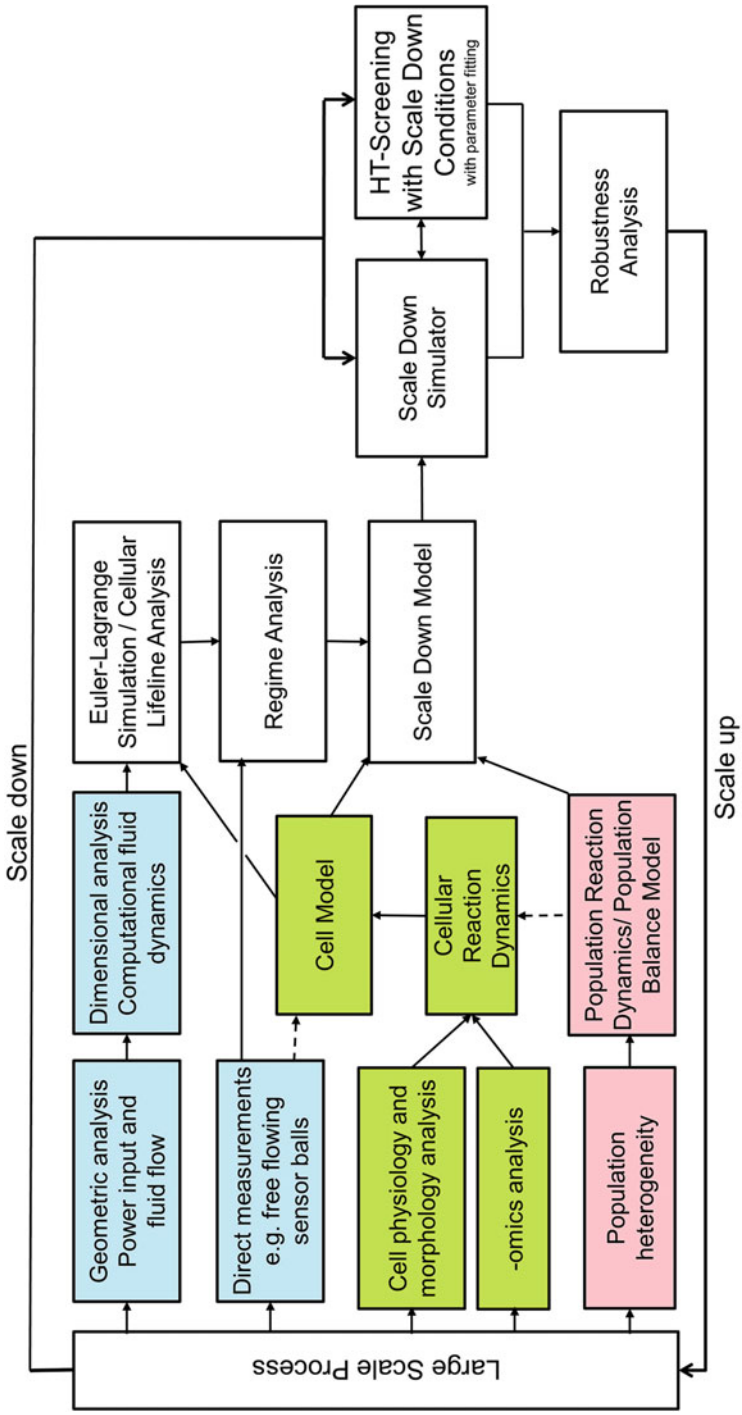


Fig. 2 Schematic outline of state-of-the-art methods for bioprocess scale-down. Scale-down starts with an analysis of the industrial large-scale bioprocess. While current approaches dominate in fluid dynamic characterization, the generation of a reliable cell model (digital twin) and the characterization of the cellular reaction dynamics, i.e. parametrization of the model, are most important for scale-down experiments. The three pillars of bioprocess characterization are highlighted in colors: reactor (blue), cell (green), populations (red)

which a defined pH value is set in each of the reactors [68]. When using plug-flow reactors, a gradient is established within the system. Then, the sampling in different positions along the plug-flow reactor also allows an insight into the time course of the cellular reactions [69, 70]. In contrast to multi-compartment systems, scale-down simulators with a pulse-based feeding are easier to establish and to run in parallel. Parameter control, however, may be more difficult to achieve due to the restriction of the feeding profiles. Since the feeding profile is easy to change (e.g., distance between the feed pulses), pulse-based systems also seem to be well applicable for robustness analyses. Alternative approaches, in which installations (e.g., plates between the different stirrers to restrict the tangential flow [71]) are realized in a laboratory reactor to extend the mixing time to the order as it is measured in the large reactor, can, in individual cases, reproduce the industrial process quite well, but are technically more complex.

4.1.1 Monitoring of the Cellular State Across Different Scales

The most successful scale-down methodology will maintain the physiological state of cells across lab and industrial scales. Naturally, it is the most suitable pre-requisite to obtain similar results, and should be considered as scaling parameter, although the examination of the physiological cell status is not easy to quantify with suitable measures. The impact of gradient formation on physiology has to be investigated with the measurement of sensitive parameters, e.g. the energy charge, stress response factors, and the respiratory activity, among others [68, 72–74]. Additionally, the physiological state may vary from cell to cell, which demands the consideration of population heterogeneity. It has been observed several times that gradient formation in fed-batch cultivation mode has an impact on population heterogeneity [75]. It adds additional parameters that lead to different phenotypes in culture (Fig. 3).

In natural habitats, mainly the cell cycle, cell ageing, and epigenetic regulation are known to have a great impact on the evolvement of phenotype diversity [77]. Stochastically asymmetric growth and mutation events drive the formation of subpopulations, which might be even better adapted to a previous change in an environment. Nevertheless, these events usually lead to lower yields in processes, which are conducted in bioreactors [78]. The role of cell cycling on the development of subpopulations in industrial bioprocesses, however, is not clear yet, while it was found out that the dominant driver for different protein concentrations, and thus various metabolic activity, is the growth rate in *Pseudomonas putida* [79].

In particular, singe-cell based and sensitive volumetric measurement techniques can provide new information about the impact of gradients on the cellular viability and metabolic activity and the formation of subpopulations independently of the scale-down system. Suitable monitoring technologies in combination with a physiological understanding of stress responses support the identification of the suitable scale-down conditions, as it puts the cell in the center of the investigation of consequences of gradient formation in the liquid phase. Such technologies, including proper accompanying off-line measurements, allow one to properly model the stress

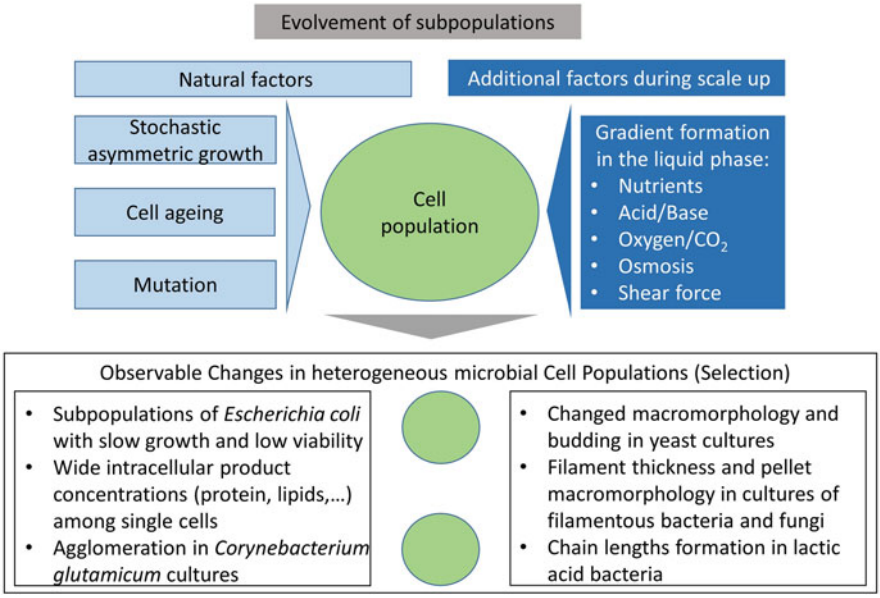


Fig. 3 Parameters that are putative effectors on the formation of subpopulations. Examples for microbial cultivations are own observations, further described in [76]

response and provide a basis for systems biology interpretation, which deepens the methodological understanding of cellular responses in the large-scale environment.

So far, investigations have shown that the exposure of cells to gradients leads to a higher population heterogeneity under scale-down conditions. This was examined in particular for protein concentrations in *E. coli* [80]. It was found that the dynamics of glycolysis might play an important role in the development of non-growing subpopulations [81]. One way to observe this evolution of subpopulations with sufficient accuracy and time-resolution is the creation of a strain with a reporter protein that can be quantified by fluorescence, which enables the application of fluorescence-assisted characterization of single cells, and eventually cell sorting [38, 76]. For example, the green fluorescence protein can be used, if coupled to automated sampling, eventually coupled with a multiplexer, and spectroscopic methods like flow cytometry, for a statistically proven detection of subpopulations [82]. In *P. putida* cultivations, the change in DNA content in individual cells was investigated under different environmental growth conditions with flow cytometric analysis at various dilution rates in chemostat experiments. The impact of oxygen deprivation, solvent exposure, and iron availability on DNA replication was also investigated [83]. The application of flow cytometry and cell staining to characterize population subgroups was described in several other studies as well [76, 84]. Nevertheless, this is a challenging technique to apply in bacteria due to their size. In bacteria, the quantification based on fluorescence is subject to genetic noise [75], which in this case might not be predominantly affected by large-scale cultivation

conditions, but constantly present across scales. The results, however, would be biased by the noise, if the samples measured are not enough to ensure a statistically valid distribution. Nevertheless, the methodology is suitable to understand the effects of gradients on cell viability and vitality while a large amount of cells can be examined in a considerably short time. While the possibility of the application of fluorescent markers represents an approach mostly on the expression level, cell staining can be used to investigate several cellular components, including the quantification of metabolite concentrations. If accumulated intracellularly, metabolite concentrations can be used to assess population heterogeneity as well, if the metabolite can be quantified with sufficient accuracy in individual cells. Flow cytometry is able to quantify the accumulation of intracellular lipids in microbial cells, and thus identify subpopulations of different lipid contents [85], e.g. with Nile Red or Bodipy® stains. The accumulation of lipids and also other components might correlate with changes in macromorphology of organisms. The measurement of lipid content with optical methods can lead to conclusions about metabolic activity of individual cells. This has been shown for heterotrophic algae, which accumulate to large extent polyunsaturated fatty acids in lipid droplets. While using light microscopy and 3-dimensional holographic microscopy, the individual lipid storage in cells was measured based on their individual cell size [86]. With rapid image analysis using trained software, image acquisition can be performed in flow cells that are connected to a cultivation. Automated workflows that offer considerably fast analysis of populations similar to flow cytometry are feasible, without the requirement of staining.

Besides intracellular product accumulation, the macromorphology can provide suitable information about the cell status and the impact of gradient formation on it. It was examined that the cell size of *S. cerevisiae* cultures changed with the degree of environmental heterogeneity in a three-compartment scale-down reactor [87]. This happened in parallel to growth reduction and side metabolite accumulation with a concomitant change of the sterol content, in comparison with homogeneous growth conditions. Cells showed a diverse macromorphology under scale-down cultivations, which supports the hypothesis that population heterogeneity is rather increased under growth in gradients. A morphologic response of cells to scale-down cultivation conditions can also lead to agglomeration due to stress response. Although the macromorphology of individual cells may stay unchanged, the secretion of side products or proteins supports the agglomeration of cells. Observations with laser-light back-reflection for cell particle size measurement indicated the formation of clumps of *C. glutamicum*, when exposed to oscillatory oxygen supply, either in a scale-down reactor concept (three-compartment reactor) or in shake flask cultures with interrupted shaking [88]. Agglomeration under fluctuating oxygen availability was postulated to be a result of increased secretion of biofilm forming metabolites, e.g. in *Mycobacterium tuberculosis* and *E. coli* cultures in connection with oxidative stress response [89, 90].

In case of filamentous organisms, macromorphological similarity across scales is often achieved only if the shear force regime is maintained. Mechanical shear forces as they appear close to the stirrer can lead to filament disruption, with consequences

on growth and secretion profiles. Up to now, however, the change of macromorphology due to scaling effects, e.g. an oscillating shear force regime, has not been investigated thoroughly. Nevertheless, it can be assumed that the macromorphology of filamentous organisms will change in comparison with the lab scale if the exposure time to high-shear forces is diminished, like it most probably is in large-scale cultivations at high cell densities and elevated viscosities, where large residence times exist in different compartments. The knowledge that exists so far about how a changed shear force regime influences the process performance [91, 92] leads to the assumption that macromorphology is an important parameter to consider while choosing a suitable scale-down system. Alternating shear forces can be achieved by interrupted stirring, which usually couples low-shear stress to oxygen limitation in stirred tank reactors, or in multi-compartment reactors, in which low-shear and high-shear regimes are applied at similar gas mass transfer rates. The application of other reactor systems beyond stirred tanks can support the investigation of consequences of low-shear forces on the macromorphology, physiology, and overall process performance as recently described for clavulanic acid production with *Streptomyces clavuligerus* in shaken bioreactors. Secretion of clavulanic acid was strongly diminished while thicker filaments were observed [93]. Consequently, morphological monitoring in an automated manner [94] is a promising technique to identify crucial characteristics for growth and product formation under specific environmental regimes.

Finally, macromorphological heterogeneity can be modeled to describe the response of a cell to environmental perturbations. In silico prediction of physiological population heterogeneity was conducted by a combination of computational fluid dynamics (CFD) and a cell cycle model of *P. putida* [95]. It was observed that 72% of the cells switched between standard and multifork replication and 52.9% showed higher than average adenosine triphosphate (ATP) maintenance demands (12.2%, up to 1.5 fold). Such an approach, however, requires sufficient knowledge of the interaction between gradient formation and consequences for the macromorphology of a population. This still represents a bottleneck as the time frequency with which morphological changes are measured might be inadequate to achieve a sufficient accuracy while correlating the response to specific regulatory events in a cell. In recent years, however, many more techniques like in situ microscopy and others are being developed rapidly. It is hoped that they become more applicable in biotechnological processes operating at elevated cell densities. The impact of gradients in the liquid phase on the formation of phenotypic heterogeneity can be investigated also if spectroscopic methods are coupled to microfluidic devices, when the growth of single cells can be monitored constantly [96]. The aforementioned methodologies will increase the possibility for the consideration of heterogeneity in population balance models and their integration in the description of consequences of gradient formation. So far, the few attempts rely on physiological measures, e.g. the adaptation to substrate excess [97, 98], but investigations will benefit from the additional consideration of macromorphological characterization data in future.

4.1.2 Combination of CFD Approaches with Mechanistic Models (Euler–Lagrange) to Describe the Large Scale

Computational fluid dynamics has been traditionally used to describe the flow under defined conditions, as a way of characterizing the heterogeneous conditions in large-scale bioreactors. With the increasing computational power it is possible to implement in such CFD approaches cellular reaction kinetics. For the first time this interaction between the intracellular state of the individual cells of the population and the turbulent flow field in the bioreactor has been realized by Lapin et al. [99]. This so-called Euler–Euler approach considers gas, liquid, and biophase as a continuum and is an answer to the very complex simulation, which is dependent on the definition and resolution of the reactor into flow cells. Later, the pioneering paper by Lapin et al. [100] was the first approach to couple a CFD model of a bioreactor with a Lagrangian approach for the combined solution of flow patterns and cellular kinetics.

In this Lagrangian–Euler approach the liquid phase is treated as a continuum (Euler) and the dispersed phase is tracked using Lagrangian representation. While this modeling approach was first used in gas–liquid simulations, here cells with their specific metabolic reaction network were described as discrete entities. With these models, individual cells are monitored with respect to their experience of local environments in relation to the fluid-dynamic distribution and pathways within the bioreactor. This kind of structured-segregated approach realizes that the individual history of a biological entity determines their reaction. With a big computational power this can be realized also in a three-dimensional turbulent field. This approach allows one to get an indication of the heterogeneity in the biotic and abiotic phases of the reactor and it considers the individual history of the cells as important for its final response. By tracking the pathway of a single particle over time it is possible to derive lifelines of a big number of cells and thus draw conclusions for regimes, i.e. conditions which should be represented in scale-down experiments.

In summary, the use of mathematical descriptions of large-scale bioreactors by the combination of Euler–Lagrangian approaches is very illustrative and has made major progress during the last years.

However, there are some limitations:

1. Due to calculation expense it is not possible to consider realistic amounts of cells. Current computational approaches consider approx. 100,000 cells, which is enough to see and follow the population dynamics. Currently this number of cells is fixed. However, it would be interesting to consider growth and an increase in cell number over time.
2. The cellular models and the parameters used in these studies are mostly derived from continuous experiments (mostly chemostats), i.e. from experimental conditions which do not reflect the large scale. Thus, as discussed above, the reactions in a real reactor may be totally different. Therefore urgent approaches and methods which describe how a cellular model can be derived and parameterized

are needed. In our opinion, there is a need to look for alternative approaches which are easily applicable and better reflect the large-scale situation.

3. Due to the very high computing power, which is needed to solve these Euler–Lagrangian models with a reasonable resolution, the current models cannot describe the process dynamics over time, but only represent a very narrow time point of the cultivation. Nevertheless, this kind of simulations, e.g. if they are performed at different time points of a cultivation can provide important information to plan scale-down experiments.
4. Current models only consider the liquid phase. The implementation of the gas phase would additionally need much computing power and in our opinion it would be very laborious to validate these models.
5. An important characteristic of living systems in connection with their adaptation to the environment is the heterogeneity in a population. Physiological (i.e., metabolic) and genetic heterogeneity ensure the survival of a population of cells if environmental changes occur (stress phenomenon) and has been described historically as the survival of the fittest. As growth in a bioreactor is related to different stresses in different phases of a process and as additional perturbations that occur in large-scale bioreactors add a further stress layer, the population heterogeneity in a bioprocess is an important feature, which needs to be monitored and can be used for a validation of similarity between different process scales.

As a consequence, every scale-down approach needs to start with a good understanding of the large scale especially in view of the cellular response dynamics. Furthermore, these cellular response dynamics must be reduced in mathematical models, the so-called digital twins, to the characteristics essential for the process scale-down. Finally, methods need to be implemented which help to validate the quality of the scale-down – and this is only possible by measurements.

4.2 Execution of Scale-Down Experiments

4.2.1 Combination of Scale-Down Experiments with Model-Based Approaches

As shown in the scale-down scheme in Fig. 1, the characterization of the larger-scale bioreactor environment is followed by transferring the environmental blueprint to the laboratory-scale simulator in which the actual scale-down experiments are executed. To achieve this transfer, the digital twin of the bioprocess should contain model units that adequately describe the physiology of the cells, as well as the cultivation process and geometric analysis of the bioreactors involved. These details should be digitally embedded into the definition of the scale-down model (Fig. 1). Thereafter, the application of the modeling framework (digital twin) in the context of scale-down experimentation can take two forms: (1) the design of the scale-down

experiments and (2) the interpretation of the data from scale-down experiments. These two branches of application are presented in the following sections, with respect to the state of the art, as reported in current literature and future perspectives.

Model-Based Design of Scale-Down Experiments

The concept of scale-down experiments developed in the past 3 decades usually involves creating one type of stress (e.g., dissolved oxygen limitation, excess substrate, excess metabolite concentrations, and perturbations in pH) in the scale-down simulator. However, considering an actual larger-scale bioreactor, concentration gradients arise from mixing effects in a 3-dimensional space, combined with the uptake of substrates and release of metabolites by cells. Moreover, the type of gradients are always coupled and may co-exist (e.g., pH-oxygen-substrate gradients) in the larger bioreactor. Therefore, at best, the scale-down simulators are only a gross estimate of the actual environments in large bioreactors. Additionally, the fraction of cells exposed to a given gradient in a large-scale bioreactor has been variable in the definition of the scale-down model. In multi-compartment scale-down simulators, this has been in the range of 10%–30%, whereas in pulse-based single-compartment simulators, the total population is subjected to the stresses, without population subgroups. Notwithstanding these challenges, important physiological responses have been reported by researchers using these physical approximations of larger bioreactors. In this light, experimental set-ups for scale-down studies can improve greatly when they are combined with the ideas of digitalization in the industry 4.0 era.

The scale-down model, defined on the blueprint of the environmental heterogeneity of the larger scale, will contain process specifications such as the gradient profiles, zone definitions with defined boundaries, residence time distributions, magnitudes of gradients, frequencies and other experimental inputs that should be implemented in the simulator. With such a scale-down model in hand, the scale-down experiment can be designed on a computer, and results of the design sent to intelligent equipment (pumps, pipettes, agitators, liquid handlers, etc.) to control a small-scale bioreactor to mimic the blueprint of the larger-scale bioreactor. That is, mixed-gradient zones, containing excess substrate with acidic pH, or limited oxygen with high hydrodynamic stress, or any desired combination of gradients can be created with the model, and the size of the zone and its frequency/duration can be varied randomly to more closely resemble actual gradient dynamics in the larger scale. A step in this direction is the work of Anane et al. [29] who used a mechanistic model of *E. coli* [101], in combination with a mechanistic description of the gradient profiles of a multi-compartment scale-down bioreactor [1] to calculate glucose pulses in high-throughput scale-down experiments. The outputs of the two models were integrated into the operation scheme of the high-throughput system to reproduce heterogenous glucose conditions in minibioreactors. The calculated glucose pulses were commensurate with the physiology of the strain, as the pulse sizes were derived from uptake capacities and physiological limits of the strain, as well as

mixing effects in the scale-down bioreactor. The authors reported significant yield losses, incorporation of non-conventional amino acids into the recombinant protein product and accumulation of metabolites in response to the calculated gradients.

Although these results were similar to observations in other non-model-based scale-down approaches [69, 102, 103], the added advantage of using the model was the flexibility of stress definition, in which one parallel experimental set-up was used to implement six different stress zones in the scale-down system, which otherwise could not be done manually.

A further advancement of this concept is the application of models to design intelligent scale-down experiments. In all the previously discussed scale-down methods, the researcher sets out a prior design space, within which the stresses and gradient profiles are pre-defined (manually or using a model) and executed during the experiment. The use of digital twins to advance scale-down design should involve designing the stresses as the experiment runs. In other words, the nature of gradients to be imposed on the culture at time point t2 will depend on its state at time point t1. This will prevent overestimating or underestimating the magnitude of gradients, as well as the exposure time of the cells to a given gradient. Such an adaptive re-design technique was employed by Cruz Bournazou and colleges in designing optimal experiments to maximize information content for model parameter identification [40, 104]. Although no scale-down efforts were made in these works, the authors demonstrated the ability to re-define feed regimes based on the current state of the culture and model predictions for a given time window. Such a model-based adaptive system can easily be employed to execute dynamic scale-down experiments.

The addition of digital twin concepts to the definition of the scale-down model offers flexibility of stress definition, automation, and a high turnover in experimental throughput, to drive the digital revolution in bioprocess engineering, as discussed by Neubauer et al. [30]. The few pioneering works published so far point in the future direction where mathematical methods, in the form of digital twins, will help to design more informative and smart (scale-down) experiments, to move away from the traditional, commonly used static design of experiment (DoE) paradigm.

Model-Based Interpretation of Scale-Down Data

Scale-down bioreactors offer important insights into cellular behavior under heterogeneous fermentation conditions of larger-scale bioreactors. The data is usually interpreted at the macroscopic level, by comparing metabolite, substrate, and growth profiles to cultivations under homogeneous conditions. In a few studies, derivative indices (e.g., specific uptake rates, yield coefficients) have been calculated from the raw scale-down data to support the interpretation of the data, e.g. in [105]. A few extensions of the data space in scale-down experiments involve molecular level analysis. For instance, Simen and co-workers used transcriptomics data from an STR-PFR scale-down bioreactor to monitor different gene expression levels under short-term and long-term substrate fluctuations in *E. coli* culture [58]. In another

study, quantitative metabolomics was used to monitor the consumption of amino acids in scale-down cultivations of *Bacillus megaterium* expressing green fluorescent protein. The metabolomics results from the scale-down conditions were then used to design a better feed composition for the process [106].

It is important to note that a strain's response to heterogenous environments in a scale-down bioreactor is the total sum of its molecular level responses. What if a particular response characteristic, such as accumulation of non-conventional amino acids or de-activation of acetate cycling (accumulation of acetate) in *E. coli* could be traced to particular metabolic pathways and mechanisms? What if data from a scale-down bioreactor could be used to trace specific metabolic fluxes of a clone as it is exposed to various concentration gradients? Such information would be useful, not only in strain engineering, but also in designing efficient processes at industrial scale. Advanced modeling of scale-down data, i.e. fitting mechanistic and dynamic metabolic models to data from scale-down cultivations can reveal specific pathways that are active under given heterogeneous conditions. This is offcourse assuming that the parameterization of the model includes such metabolic and physiological indices. The flux terms to be fitted to the scale-down data should be an integral part of the building of the cell model (Fig. 1). Again, Anane et al. [29] fitted data from parallel scale-down cultivations of *E. coli* under multiple glucose gradient conditions to a mechanistic model describing the process and the strain. The authors found that the different responses of the strain to the gradients translated directly into different values of the parameters of the model. Therefore, the model expanded on the primary data of the scale-down experiment, and expanded the interpretation of the available data for better process and strain design. In a similar study, Janakiraman and co-workers used multi-variate data analysis techniques to interpret scale-down data [107]. Their aim was to establish comparability between scale-down cultivations in Ambr15® minibioreactors and cultivations in 15,000 L manufacturing scale, by applying principal component analysis to the scale-down datasets. By employing this model-based approach, they were able to clearly identify that the runs in both scales were statistically similar to each other, a conclusion that would have been difficult to draw by looking at the raw scale-down data.

For digital twins to be applicable in this sense, there are a few pre-requisites the model of the bioprocess must fulfill: (1) the parameter estimates in the cell model must be subjected to rigorous validity and uncertainty tests, as presented by Anane et al. [18]. The reported parameter values should always be accompanied by confidence intervals at valid significance levels, to be able to derive biological meaning from the model results. (2) The model should be just as detailed as is necessary for its application. As pointed out by Gábor and Banga [108], the parsimony principle should always be applied in building the model: i.e., the number of parameters should not be more than those required to describe the process in its simplest form [108, 109], and (3) the model should be constantly updated to include the most recent research findings in cell physiology and metabolism. The physiological accuracy of the model should be ascertained by subject matter experts in the field, which may not necessarily be the modeler (mathematician).

4.2.2 High-Throughput Execution of Scale-Down Experiments in Parallel Cultivation Systems

High-throughput experiments in parallel cultivation platforms have become common in bioprocess development laboratories. In the past decade, there has been an exponential increase in the adoption of these systems for early bioprocess development [45, 110, 111]. At the same time, due to Quality-by-Design (QbD) guidelines, there has been an increasing demand to fully characterize bioprocesses at the development phase, to forestall unforeseen consequences of the final process, upon scale-up [112, 113]. This requirement demands that all conditions, including actual large-scale process conditions are considered and tested in the early development phases of the process. Therefore, the question of whether cultivations in minibioreactors are adaptable to mimic concentration gradients and the heterogeneous environments that exist in large-scale bioreactors has become very important, and should be addressed.

A few studies conducted in high-throughput cultivation systems that consider the heterogeneous conditions of larger bioreactors are reported in the literature. As described above, Janakiraman et al. [107] matched the volumetric aeration rates (vvm) between parallel Ambr15® cultivations of CHO cells and a 15,000 L production-scale bioreactor. They used this criterion to mimic the carbon dioxide profile of the production bioreactor in the minibioreactor cultivations, which led to similar productivity and product quality profiles in both the 15 ml bioreactors and the 15,000 L scale. In another study, Velez-Suberbie et al. [114] used the power per unit volume (P/V) as a scale-down criterion to compare Ambr15 cultivations of *E. coli* with 20 L bioreactor cultivations [114]. Perhaps the most comprehensive work in this regard was reported by Anane et al. [29], who used model-calculated glucose pulses to induce both dissolved oxygen and glucose gradient zones in 15 ml parallel minibioreactors. A key aspect of their work was the use of robotic liquid handling stations and mechanistic models in the operation of the scale-down set-up. These smart equipment were interphased with the minibioreactors, such that model outputs describing specific gradient conditions could be implemented in selected minibioreactors by the robotic system. Their results in *E. coli* fermentation development showed significant accumulation of non-conventional amino acids in the recombinant protein product, as well as accumulation of acetate in the scale-down cultivations, when compared to cultivations under homogeneous conditions.

The results of scale-down cultivations as performed in high-throughput minibioreactor systems so far show that it is possible to mimic large-scale environmental conditions in miniaturized bioreactors. Particularly, the physiological responses of both *E. coli* and CHO cells to the induced heterogeneous conditions in minibioreactors, as discussed above, is a proof of concept that gradient profiles that are relevant in industrial-scale cultivations can be reproduced in milliliter scale for scale-down studies. However, the adoption of enabling technological methods, such as robotic liquid handling stations and mechanistic modeling is fundamental for the successful operation of such minibioreactor facilities as scale-down platforms.

The adoption of such parallel cultivation systems and their combination with robotic liquid handling stations will ensure that a large number of gradient profiles, defined in the scale-down model, can be tested in a single parallel run. Additionally, such high-throughput systems can be used for strain screening under conditions that are amenable to the larger scale, to select the most robust strain for further development and scale-up.

5 General Conclusions and Perspectives

The lead times of biotechnological products, especially biopharmaceuticals, from discovery to market, can be up to 15 years [30]. Although other issues such as clinical trials may contribute to this time, bioprocess development and troubleshooting scale-up problems are key contributors to the lengthy lead times. The use of parallel cultivation systems and robotics has, undoubtedly, reduced these process development times significantly [45, 110]. Prior to screening, the development of strains is nowadays performed in a high-throughput manner, e.g. with the use of standardized genetic methods [115] and non-targeted high-throughput strain engineering [116]. Thus, the bottleneck of a faster overall bioprocess development is shifted from strain engineering to screening and cultivation development. The use of parallelized minibioreactor systems for both screening and upstream process development, as demonstrated in different studies [29, 107, 114], will greatly relieve this bottleneck, and ensure that a potential bioprocess reaches production within the earliest possible times. Additionally, the framework of screening under scale-down conditions and the associated methods will not only facilitate rapid bioprocess development, but also ensure a consistent and efficient cultivation process development by taking into account all the possible cultivation conditions that would be encountered upon process scale-up.

Digital twins have become an integral part of bioprocess development and process control. Particularly, the high degree of parallelization and automation of the development process, the integration of PAT and the requirements for a higher robustness of the processes in connection with an improved process control could only be realized through the comprehensive implementation of mathematical and statistical methods. Thus, the current challenge lies especially in the fusion of the individual tools into a uniform overall system.

The application of new possibilities that arose from the ongoing development of sensor technology and the corresponding data processing allows a stronger consideration of cell-to-cell variation and cellular features as scaling parameters. Such technologies, including proper accompanying off-line measurements, allow one to properly model stress responses and provide a basis for the integration of systems biology knowledge to deepen the methodological understanding of cellular responses in a large-scale environment. This can support the identification of suitable scale-down systems with the cell status as scaling factor as it represents the central

location for the product synthesis. It fosters the application of population balances and their integration into model-based descriptions of scale-up effects.

References

1. Anane E, Sawatzki A, Neubauer P, Cruz Bournazou MN (2019) Modelling concentration gradients in fed-batch cultivations of *E. coli* – towards the flexible design of scale-down experiments. J Chem Technol Biotechnol 94:516–526
2. Neubauer P, Cruz N, Glauche F, Junne S, Knepper A, Raven M (2013) Consistent development of bioprocesses from microliter cultures to the industrial scale. Eng Life Sci 13:224–238
3. Neubauer P, Junne S (2016) Scale-up and scale-down methodologies for bioreactors. In: Mandenius CF (ed) Bioreactors: design, operation and novel applications. Wiley-VCH Verlag GmbH, Weinheim, pp 323–354
4. Reitz C, Fan Q, Neubauer P (2018) Synthesis of non-canonical branched-chain amino acids in *Escherichia coli* and approaches to avoid their incorporation into recombinant proteins. Curr Opin Biotechnol 53:248–253
5. Wang G, Haringa C, Tang W, Noorman H, Chu J, Zhuang Y, Zhang S (2020) Coupled metabolic-hydrodynamic modeling enabling rational scale-up of industrial bioprocesses. Biotechnol Bioeng 117:844–867
6. Grieves M, Vickers J (2016) Digital twin: mitigating unpredictable, undesirable emergent behavior in complex systems. In: Transdisciplinary perspectives on complex systems: new findings and approaches, pp 85–113
7. Grossmann I (2005) Enterprise-wide optimization: a new frontier in process systems engineering. AIChE J:1846–1857
8. Qi Q, Tao F (2018) Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison. IEEE Access 6:3585–3593
9. Batstone DJ, Keller J, Angelidaki I, Kalyuzhnyi SV, Pavlostathis SG, Rozzi A, Sanders WTM, Siegrist H, Vavilin VA (2002) The IWA anaerobic digestion model no 1(ADM 1). Water Sci Technol 45:65–73
10. Tsugawa H (2018) Advances in computational metabolomics and databases deepen the understanding of metabolisms. Curr Opin Biotechnol 54:10
11. Kitano H (2002) Computational systems biology. Nature 420:206–210
12. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Review of cellular metabolism. In: Metabolic engineering. Academic Press, San Diego, pp 21–79
13. Varma A, Palsson BO (1994) Metabolic flux balancing: basic concepts, scientific and practical use. Bio/Technology 12:994–998
14. Marchisio MA, Stelling J (2009) Computational design tools for synthetic biology. Curr Opin Biotechnol 20:479–485
15. Saeys Y, Inza I, Larrañaga P (2007) A review of feature selection techniques in bioinformatics. Bioinformatics 23:2507–2517
16. Bailey JE (1998) Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Prog 14:8–20
17. Koutinas M, Kiparissides A, Pistikopoulos EN, Mantalaris A (2012) Bioprocess systems engineering: transferring traditional process engineering principles to industrial biotechnology. Comput Struct Biotechnol J 3:e201210022
18. Anane E, López CDC, Barz T, Sin G, Gernaey KV, Neubauer P, Cruz Bournazou MN (2019) Output uncertainty of dynamic growth models: effect of uncertain parameter estimates on model reliability. Biochem Eng J 150:107247

19. Muñoz-Tamayo R, Puillet L, Daniel JB, Sauvant D, Martin O, Taghipoor M, Blavy P (2018) Review: to be or not to be an identifiable model. Is this a relevant question in animal science modelling? Animal 12:701–712
20. Villaverde AF, Barreiro A, Papachristodoulou A (2016) Structural identifiability of dynamic systems biology models. PLoS Comput Biol 12:1–22
21. Brubaker TA (1979) Nonlinear parameter estimation. Anal Chem 51:1385A
22. Brun R, Kühni M, Siegrist H, Gujer W, Reichert P (2002) Practical identifiability of ASM2d parameters – systematic selection and tuning of parameter subsets. Water Res 36:4113–4127
23. Kravaris C, Hahn J, Chu Y (2013) Advances and selected recent developments in state and parameter estimation. Comput Chem Eng 51:111–123
24. Vajda S, Rabitz H, Walter E, Lecourtier Y (1989) Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. Chem Eng Commun 83:191–219
25. Bellman R, Astrom KJ (1970) On structural identifiability. Math Biosci 7:329–339
26. Cobelli C, DiStefano JJ (1980) Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. Am J Phys 239:R7–R24
27. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J (2009) Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics 25:1923–1929
28. Neubauer P, Cruz-Bournazou MN (2017) Continuous bioprocess development: methods for control and characterization of the biological system. In: Subramanian G (ed) Continuous biomanufacturing – innovative technologies and methods. Wiley, Hoboken, pp 1–30
29. Anane E, García ÁC, Haby B, Hans S, Krausch N, Krewinkel M, Hauptmann P, Neubauer P, Cruz Bournazou MN (2019) A model-based framework for parallel scale-down fed-batch cultivations in mini-bioreactors for accelerated phenotyping. Biotechnol Bioeng 116:2906–2918
30. Neubauer P, Glauche F, Cruz-Bournazou MN (2017) Editorial: bioprocess development in the era of digitalization. Eng Life Sci 17:1140–1141
31. Narayanan H, Luna MF, von Stosch M, Cruz Bournazou MN, Polotti G, Morbidelli M, Butté A, Sokolov M (2020) Bioprocessing in the digital age: the role of process models. Biotechnol J 15:1–10
32. Noorman H (2011) An industrial perspective on bioreactor scale-down: what we can learn from combined large-scale bioprocess and model fluid studies. Biotechnol J 6:934–943
33. Petsagkourakis P, Sandoval IO, Bradford E, Zhang D, del Rio-Chanona EA (2020) Reinforcement learning for batch bioprocess optimization. Comput Chem Eng 133:106649
34. Mandenius CF, Brundin A (2008) Bioprocess optimization using design-of-experiments methodology. Biotechnol Prog 24:1191–1203
35. Wechselberger P, Sagmeister P, Herwig C (2013) Model-based analysis on the extractability of information from data in dynamic fed-batch experiments. Biotechnol Prog 29:285–296
36. Covert MW, Xiao N, Chen TJ, Karr JR (2008) Integrating metabolic, transcriptional regulatory and signal transduction models in *Escherichia coli*. Bioinformatics 24:2044–2050
37. Haringa C, Deshmukh AT, Mudde RF, Noorman HJ (2017) Euler-Lagrange analysis towards representative down-scaling of a 22 m^3 aerobic *S. cerevisiae* fermentation. Chem Eng Sci 170:653–669
38. Delvigne F, Goffin P (2014) Microbial heterogeneity affects bioprocess robustness: dynamic single-cell analysis contributes to understanding of microbial populations. Biotechnol J 9:61–72
39. Barz T, Lopez Cardenas DC, Cruz Bournazou MN, Körkel S, Walter SF (2016) Real-time adaptive input design for the determination of competitive adsorption isotherms in liquid chromatography. Comput Chem Eng 94:104–116
40. Cruz Bournazou MN, Barz T, Nickel DB, Lopez Cárdenas DC, Glauche F, Knepper A, Neubauer P (2017) Online optimal experimental re-design in robotic parallel fed-batch cultivation facilities. Biotechnol Bioeng 114:610–619

41. Dörr M, Fibinger MPC, Last D, Schmidt S, Santos-Aberturas J, Böttcher D, Hummel A, Vickers C, Voss M, Bornscheuer UT (2016) Fully automatized high-throughput enzyme library screening using a robotic platform. Biotechnol Bioeng 113:1421–1432
42. Haby B, Hans S, Anane E, Sawatzki A, Krausch N, Neubauer P, Cruz Bournazou MN (2019) Integrated robotic mini bioreactor platform for automated, parallel microbial cultivation with online data handling and process control. SLAS Technol 24:569–582
43. Unthan S, Radek A, Wiechert W, Oldiges M, Noack S (2015) Bioprocess automation on a mini pilot plant enables fast quantitative microbial phenotyping. Microb Cell Factories 14:32
44. Lattermann C, Büchs J (2015) Microscale and miniscale fermentation and screening. Curr Opin Biotechnol 35:1–6
45. Hemmerich J, Noack S, Wiechert W, Oldiges M (2018) Microbioreactor systems for accelerated bioprocess development. Biotechnol J 13:1–9
46. Lara AR, Galindo E, Ramírez OT, Palomares LA (2006) Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. Mol Biotechnol 34:355–382
47. Oldshue JY (1966) Fermentation mixing scale-up techniques. Biotechnol Bioeng 8:3–24
48. Oosterhuis NMG (1984) Scale-up of bioreactors. TU Delft 162
49. Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E, Schweder T, Hamer G, O'Beirne D, Noisommit-Rizzi N, Reuss M, Boone L, Hewitt C, McFarlane C, Nienow A, Kovacs T, Trägårdh C, Fuchs L, Revstedt J, Friberg PC, Hjertager B, Blomsten G, Skogman H, Hjort S, Hoeks F, Lin HY, Neubauer P, Van der Lans R, Luyben K, Vrabel P, Manelius A, Manelius Å (2001) Physiological responses to mixing in large scale bioreactors. J Biotechnol 85:175–185
50. Larsson G, Törnkvist M, Ståhl Wernersson E, Trägårdh C, Noorman H, Enfors SO (1996) Substrate gradients in bioreactors: origin and consequences. Bioprocess Eng 14:281–289
51. Brand E, Junne S, Anane E, Cruz-Bournazou MN, Neubauer P (2018) Importance of the cultivation history for the response of *Escherichia coli* to oscillations in scale-down experiments. Bioprocess Biosyst Eng 41:1305–1313
52. Sweere APJ, Luyben KCAM, Kossen NWF (1987) Regime analysis and scale-down: tools to investigate the performance of bioreactors. Enzym Microb Technol 9:386–398
53. Lin HY, Mathiszik B, Xu B, Enfors SO, Neubauer P (2001) Determination of the maximum specific uptake capacities for glucose and oxygen in glucose-limited fed-batch cultivations of *Escherichia coli*. Biotechnol Bioeng 73:347–357
54. Bylund F, Collet E, Larsson G, Enfors SO, Larsson G (1998) Substrate gradient formation in the large-scale bioreactor lowers cell yield and increases by-product formation. Bioprocess Eng 18:171–180
55. Neubauer P, Häggström L, Enfors SO (1995) Influence of substrate oscillations on acetate formation and growth yield in *Escherichia coli* glucose limited fed-batch cultivations. Biotechnol Bioeng 47:139–146
56. Xu B, Jahic M, Blomsten G, Enfors SO (1999) Glucose overflow metabolism and mixed-acid fermentation in aerobic large-scale fed-batch processes with *Escherichia coli*. Appl Microbiol Biotechnol 51:564–571
57. Nienow AW (2006) Reactor engineering in large scale animal cell culture. Cytotechnology 50:9–33
58. Simen JD, Löffler M, Jäger G, Schäferhoff K, Freund A, Matthes J, Müller J, Takors R, Feuer R, von Wulffen J, Lischke J, Ederer M, Knies D, Kunz S, Sawodny O, Riess O, Sprenger G, Trachtmann N, Nieß A, Broicher A (2017) Transcriptional response of *Escherichia coli* to ammonia and glucose fluctuations. Microb Biotechnol 10:858–872
59. Spann R, Glibstrup J, Pellicer-Alborch K, Junne S, Neubauer P, Roca C, Kold D, Lantz AE, Sin G, Gernaey KV, Krühne U (2019) CFD predicted pH gradients in lactic acid bacteria cultivations. Biotechnol Bioeng 116:769–780

60. Paul K, Böttinger K, Mitic BM, Scherfler G, Posch C, Behrens D, Huber CG, Herwig C (2020) Development, characterization, and application of a 2-compartment system to investigate the impact of pH inhomogeneities in large-scale CHO-based processes. Eng Life Sci 20:368–378
61. Paul K, Hartmann T, Posch C, Behrens D, Herwig C (2020) Investigation of cell line specific responses to pH inhomogeneity and consequences for process design. Eng Life Sci 20:412–421
62. Buchholz J, Graf M, Freund A, Busche T, Kalinowski J, Blombach B, Takors R (2014) CO 2/HCO 3− perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in *Corynebacterium glutamicum*. Appl Microbiol Biotechnol 98:8563–8572
63. Spadiut O, Rittmann S, Dietzsch C, Herwig C (2013) Dynamic process conditions in bioprocess development. Eng Life Sci 13:88–101
64. Limberg MH, Joachim M, Klein B, Wiechert W, Oldiges M (2017) pH fluctuations imperil the robustness of *C. glutamicum* to short term oxygen limitation. J Biotechnol 259:248–260
65. Xu S, Jiang R, Mueller R, Hoesli N, Kretz T, Bowers J, Chen H (2018) Probing lactate metabolism variations in large-scale bioreactors. Biotechnol Prog 34:756–766
66. Brunner M, Doppler P, Klein T, Herwig C, Fricke J (2018) Elevated pCO2 affects the lactate metabolic shift in CHO cell culture processes. Eng Life Sci 18:204–214
67. Delvigne F, Noorman H (2017) Scale-up/scale-down of microbial bioprocesses: a modern light on an old issue. Microb Biotechnol 10:685–687
68. Cortés JT, Flores N, Bolívar F, Lara AR, Ramírez OT (2016) Physiological effects of pH gradients on *Escherichia coli* during plasmid DNA production. Biotechnol Bioeng 113:598–611
69. Junne S, Klingner A, Kabisch J, Schweder T, Neubauer P (2011) A two-compartment bioreactor system made of commercial parts for bioprocess scale-down studies: impact of oscillations on *Bacillus subtilis* fed-batch cultivations. Biotechnol J 6:1009–1017
70. Käß F, Hariskos I, Michel A, Brandt HJ, Spann R, Junne S, Wiechert W, Neubauer P, Oldiges M (2014) Assessment of robustness against dissolved oxygen/substrate oscillations for *C. glutamicum* DM1933 in two-compartment bioreactor. Bioprocess Biosyst Eng 37:1151–1162
71. Schilling BM, Pfefferle W, Bachmann B, Leuchtenberger W, Deckwer W-DD (1999) A special reactor design for investigations of mixing time effects in a scaled-down industrial L-lysine fed-batch fermentation process. Biotechnol Bioeng 64:599–606
72. Delvigne F, Boxus M, Ingels S, Thonart P (2009) Bioreactor mixing efficiency modulates the activity of a *prpoS::GFP* reporter gene in *E. coli*. Microb Cell Fact 8:15
73. Junne S, Neubauer P (2018) How scalable and suitable are single-use bioreactors? Curr Opin Biotechnol 53:240–247
74. Löffler M, Simen JD, Jäger G, Schäferhoff K, Freund A, Takors R (2016) Engineering *E. coli* for large-scale production – strategies considering ATP expenses and transcriptional responses. Metab Eng 38:73–85
75. Delvigne F, Baert J, Sassi H, Fickers P, Grünberger A, Dusny C (2017) Taking control over microbial populations: current approaches for exploiting biological noise in bioprocesses. Biotechnol J 12:1600549
76. Lemoine A, Delvigne F, Bockisch A, Neubauer P, Junne S (2017) Tools for the determination of population heterogeneity caused by inhomogeneous cultivation conditions. J Biotechnol 251:84–93
77. Avery SV (2006) Microbial cell individuality and the underlying sources of heterogeneity. Nat Rev Microbiol 4:577–587
78. Binder D, Drepper T, Jaeger KE, Delvigne F, Wiechert W, Kohlheyer D, Grünberger A (2017) Homogenizing bacterial cell factories: analysis and engineering of phenotypic heterogeneity. Metab Eng 42:145–156

79. Lieder S, Jahn M, Seifert J, von Bergen M, Müller S, Takors R (2014) Subpopulation-proteomics reveal growth rate, but not cell cycling, as a major impact on protein composition in Pseudomonas putida KT2440. AMB Express 4:1–10
80. Delvigne F, Zune Q, Lara AR, Al-Soud W, Sørensen SJ (2014) Metabolic variability in bioprocessing: implications of microbial phenotypic heterogeneity. Trends Biotechnol 32:608–616
81. Van Heerden JH, Wortel MT, Bruggeman FJ, Heijnen JJ, Bollen YJM, Planqué R, Hulshof J, O'Toole TG, Wahl SA, Teusink B (2014) Lost in transition: start-up of glycolysis yields subpopulations of nongrowing cells. Science 343:1245114
82. Brognaux A, Han S, Sørensen SJ, Lebeau F, Thonart P, Delvigne F (2013) A low-cost, multiplexable, automated flow cytometry procedure for the characterization of microbial stress dynamics in bioreactors. Microb Cell Fact 12
83. Lieder S, Jahn M, Koepff J, Muller S, Takors R (2016) Environmental stress speeds up DNA replication in *Pseudomonas putida* in chemostat cultivations. Biotechnol J 11:155–163
84. Hewitt CJ, von Caron GN, Axelsson B, McFarlane CM, Nienow AW (2000) Studies related to the scale-up of high-cell-density *E. coli* fed-batch fermentations using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. Biotechnol Bioeng 70:381–390
85. Patel A, Antonopoulou I, Enman J, Rova U, Christakopoulos P, Matsakas L (2019) Lipids detection and quantification in oleaginous microorganisms: an overview of the current state of the art. BMC Chem Eng 1
86. Marbà-Ardébol AM, Emmerich J, Neubauer P, Junne S (2017) Single-cell-based monitoring of fatty acid accumulation in *Crypthecodinium cohnii* with three-dimensional holographic and in situ microscopy. Process Biochem 52:223–232
87. Marbà-Ardébol AM, Bockisch A, Neubauer P, Junne S (2018) Sterol synthesis and cell size distribution under oscillatory growth conditions in *Saccharomyces cerevisiae* scale-down cultivations. Yeast 35:213–223
88. Lemoine A, Limberg MHMH, Kästner S, Oldiges M, Neubauer P, Junne S (2016) Performance loss of *Corynebacterium glutamicum* cultivations under scale-down conditions using complex media. Eng Life Sci 16:620–632
89. Nachin L, Nannmark U, Nyström T (2005) Differential roles of the universal stress proteins of *Escherichia coli* in oxidative stress resistance, adhesion, and motility. J Bacteriol 187:6265–6272
90. Trivedi A, Mavi PS, Bhatt D, Kumar A (2016) Thiol reductive stress induces cellulose-anchored biofilm formation in *Mycobacterium tuberculosis*. Nat Commun 7
91. Kurt T, Marbà-Ardébol AM, Turan Z, Neubauer P, Junne S, Meyer V (2018) Rocking Aspergillus: morphology-controlled cultivation of *Aspergillus niger* in a wave-mixed bioreactor for the production of secondary metabolites. Microb Cell Factories 17:128
92. Lin PJ, Scholz A, Krull R (2010) Effect of volumetric power input by aeration and agitation on pellet morphology and product formation of *Aspergillus niger*. Biochem Eng J 49:213–220
93. Gómez-Ríos D, Junne S, Neubauer P, Ochoa S, Ríos-Estepa R, Ramírez-Malule H (2019) Characterization of the metabolic response of *Streptomyces clavuligerus* to shear stress in stirred tanks and single-use 2D rocking motion bioreactors for clavulanic acid production. Antibiotics 8
94. Hardy N, Moreaud M, Guillaume D, Augier F, Nienow A, Béal C, Chaabane FB (2017) Advanced digital image analysis method dedicated to the characterization of the morphology of filamentous fungus. J Microsc 266:126–140
95. Kuschel M, Siebler F, Takors R (2017) Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. Bioengineering 4:27
96. Ladner T, Grünberger A, Probst C, Kohlheyer D, Büchs J, Delvigne F (2017) Application of mini- and micro-bioreactors for microbial bioprocesses. In: Current developments in biotechnology and bioengineering: bioprocesses, bioreactors and controls. Elsevier, Amsterdam, pp 433–461

97. Morchain J, Gabelle J-C, Cockx A (2013) Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors. AICHE J 59:369–379
98. Pigou M, Morchain J (2015) Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. Chem Eng Sci 126:267–282
99. Lapin A, Müller D, Reuss M (2004) Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler−Lagrange methods: traveling along the lifelines of single cells. Ind Eng Chem Res 43:4647–4656
100. Lapin A, Klann M, Reuss M (2010) Multi-scale spatio-temporal modeling: lifelines of microorganisms in bioreactors and tracking molecules in cells. Adv Biochem Eng Biotechnol 121:23–43
101. Anane E, López DC, Neubauer P, Cruz Bournazou MN (2017) Modelling overflow metabolism in *Escherichia coli* by acetate cycling. Biochem Eng J 125:23–30
102. Lara AR, Taymaz-Nikerel H, Mashego MR, Van Gulik WM, Heijnen JJ, Ramirez OT, van Winden WA, Van Gulik WM, Heijnen JJ, Van Winden WA, Ramírez OT, van Winden WA (2009) Fast dynamic response of the fermentative metabolism of *Escherichia coli* to aerobic and anaerobic glucose pulses. Biotechnol Bioeng 104:1153–1161
103. Soini J, Ukkonen K, Neubauer P (2011) Accumulation of amino acids deriving from pyruvate in *Escherichia coli* W3110 during fed-batch cultivation in a two-compartment scale-down bioreactor. Adv Biosci Biotechnol 02:336–339
104. Barz T, Sommer A, Wilms T, Neubauer P, Cruz Bournazou MN (2018) Adaptive optimal operation of a parallel robotic liquid handling station. IFAC-PapersOnLine 51:765–770
105. Lemoine A, Martnez-Iturralde NM, Spann R, Neubauer P, Junne S (2015) Response of Corynebacterium glutamicum exposed to oscillating cultivation conditions in a two- and a novel three-compartment scale-down bioreactor. Biotechnol Bioeng 112:1220–1231
106. Korneli C, Bolten CJ, Godard T, Franco-Lara E, Wittmann C, Universita T (2012) Debottlenecking recombinant protein production in Bacillus megaterium under large-scale conditions—targeted precursor feeding designed from metabolomics. Biotechnol Bioeng 109:1538–1550
107. Janakiraman V, Kwiatkowski C, Kshirsagar R, Ryll T, Huang YM (2015) Application of high-throughput mini-bioreactor system for systematic scale-down modeling, process characterization, and control strategy development. Biotechnol Prog 31:1623–1632
108. Gábor A, Banga JR (2015) Robust and efficient parameter estimation in dynamic models of biological systems. BMC Syst Biol 9:74
109. Rollié S, Mangold M, Sundmacher K (2012) Designing biological systems: systems engineering meets synthetic biology. Chem Eng Sci 69:1–29
110. Bareither R, Pollard D (2011) A review of advanced small-scale parallel bioreactor technology for accelerated process development: current state and future need. Biotechnol Prog 27:2–14
111. Rameez S, Mostafa SS, Miller C, Shukla AA (2014) High-throughput miniaturized bioreactors for cell culture process development: reproducibility, scalability, and control. Biotechnol Prog 30:718–727
112. Herwig C, Garcia-Aponte OF, Golabgir A, Rathore AS (2015) Knowledge management in the QbD paradigm: manufacturing of biotech therapeutics. Trends Biotechnol 33:381–387
113. Rathore AS (2009) Roadmap for implementation of quality by design (QbD) for biotechnology products. Trends Biotechnol 27:546–553
114. Velez-Suberbie ML, Betts JPJ, Walker KL, Robinson C, Zoro B, Keshavarz-Moore E (2017) High-throughput automated microbial bioreactor system used for clone selection and rapid scale-down process optimization. Biotechnol Prog 15:1–11
115. de Lorenzo V, Schmidt M (2018) Biological standards for the knowledge-based BioEconomy: what is at stake. New Biotechnol 40:170–180
116. Schallmey M, Frunzke J, Eggeling L, Marienhagen J (2014) Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. Curr Opin Biotechnol 26:148–154

Adv Biochem Eng Biotechnol (2021) 177: 29–62
https://doi.org/10.1007/10_2020_136

Published online: 15 August 2020

Digital Twins and Their Role in Model-Assisted Design of Experiments

Kim B. Kuchemüller, Ralf Pörtner, and Johannes Möller

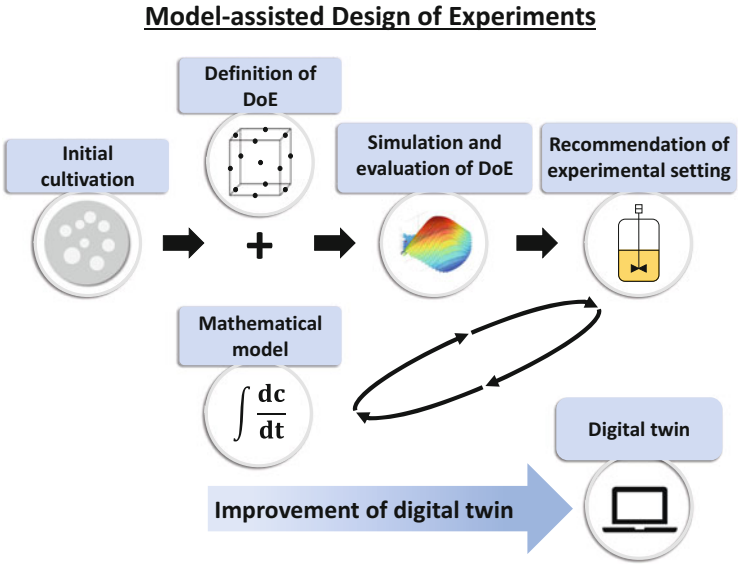
Contents

Abstract Rising demands for biopharmaceuticals and the need to reduce manufacturing costs increase the pressure to develop productive and efficient bioprocesses. Among others, a major hurdle during process development and optimization studies is the huge experimental effort in conventional design of experiments (DoE) methods. As being an explorative approach, DoE requires extensive expert knowledge about the investigated factors and their boundary values and often leads to multiple rounds of time-consuming and costly experiments. The combination of DoE with a virtual representation of the bioprocess, called digital twin, in

K. B. Kuchemüller, R. Pörtner, and J. Möller (✉)
Institute of Bioprocess and Biosystems Engineering, Hamburg University of Technology, Hamburg, Germany
e-mail: johannes.moeller@tuhh.de

model-assisted DoE (mDoE) can be used as an alternative to decrease the number of experiments significantly. mDoE enables a knowledge-driven bioprocess development including the definition of a mathematical process model in the early development stages. In this chapter, digital twins and their role in mDoE are discussed. First, statistical DoE methods are introduced as the basis of mDoE. Second, the combination of a mathematical process model and DoE into mDoE is examined. This includes mathematical model structures and a selection scheme for the choice of DoE designs. Finally, the application of mDoE is discussed in a case study for the medium optimization in an antibody-producing Chinese hamster ovary cell culture process.

Graphical Abstract



Keywords Cell culture, Experimental design, Fed-batch strategy, Process design and optimization, Quality by design

Abbreviations

A	Average
Amm	Ammonium
ANOVA	Analysis of variance
BBD	Box-Behnken design
CCC	Central composite circumscribed
CCD	Central composite designs

CCF	Central composite face centered
CCI	Central composite inscribed
CHO	Chinese hamster ovary
D	Determinant
DoE	Design of experiments
E	Eigenvalue
G	Global
Glc	Glucose
Gln	Glutamine
GMP	Good Manufacturing Practice
I	Variance
Lac	Lactate
LHSD	Latin hypercube sampling design
mAb	Antibody
max	Maximum
MBDoE	Model-based design of experiments
mDoE	Model-assisted design of experiments
min	Minimum
PAT	Process analytical technology
QbD	Quality by design
VPA	Valproic acid

Nomenclature

α	Distance to center point (-)
β_i	Unknown constants (-)
ε_i	Random error (-)
γ	Constant antibody production rate (mg cell^{-1} h^{-1})
μ	Cell-specific growth rate (h^{-1})
$\mu_{d,max}$	Maximum death rate (h^{-1})
$\mu_{d,min}$	Minimum death rate (h^{-1})
μ_{max}	Maximum growth rate (h^{-1})
c_i	Concentration of component i (mmol L^{-1})
d_i	Desirability function (−)
D	Overall desirability function (−)
i	Index (−)
k	Factors (−)
k_{Lys}	Cell lysis constant (h^{-1})
$K_{S,i}$	Monod kinetic constant for component i (mmol L^{-1})
L_i	Lower acceptable response (−)
n	Steps (−)
q_{Amm}	Ammonium formation rate (mmol cell^{-1} h^{-1})
q_{Glc}	Glucose formation rate (mmol cell^{-1} h^{-1})
q_{Gln}	Glutamine formation rate (mmol cell^{-1} h^{-1})

$q_{i,max}$	Maximum uptake rate of component i (mmol cell^{-1} h^{-1})
q_{Lac}	Lactate formation rate (mmol cell^{-1} h^{-1})
$q_{Lac,uptake}$	Uptake rate of lactate (mmol cell^{-1} h^{-1})
$q_{Lac,uptake,max}$	Maximum uptake rate of lactate (mmol cell^{-1} h^{-1})
q_{mAb}	Antibody formation rate (mmol cell^{-1} h^{-1})
R^2	Coefficient of determination (−)
U_i	Upper acceptable response (−)
x_i	Independent variables (−)
X_t	Total cell density (cells mL^{-1})
X_v	Viable cell density (cells mL^{-1})
V_i	Viability (−)
$Y_{Amm/Gln}$	Yield coefficient of ammonium formation to glutamine uptake (−)
y_i	Response (−)
$Y_{Lac/Glc}$	Yield coefficient of lactate formation to glucose uptake (−)

1 Introduction

The demand for highly effective pharmaceuticals has risen continuously over the past decades [1, 2]. From 2015 to 2018, 129 different biopharmaceuticals have been approved by the EU and the US government, representing the highest number of approvals in a 4-year period since the first biopharmaceuticals were introduced in the end of the twentieth century [3]. In 2018, a total of 374 approved biopharmaceuticals were available, including 316 with different individual active ingredients and current active registrations [4]. Trends for the future indicate a growing market share of up to 50% of the top 100 pharmaceuticals to be bio-based [5], predominantly monoclonal antibody-derived medicinal substances, followed by hormones and blood-related drugs [4]. Simultaneously, the development costs of biopharmaceuticals have increased drastically (620% from 1980 to 2013) [6]. As a result, processes become more complex and intensified, which is further increased by, e.g., changing from simple batch to more complex fed-batch or perfusion processes. The number of process variables to be monitored and their complexity have also increased. Finally, the requirements for quality management and documentation (good manufacturing practice – GMP) have also increased to guarantee quality [7]. For the design of novel bioprocesses, the process analytical technology (PAT) initiative and quality by design (QbD) philosophy require an improved understanding of the drug manufacturing processes [8].

Statistical design of experiments (DoE) methods have become common practice in process development within QbD [9]. However, induced by the explorative approach of DoE, the selection of the experimental design as well as the definition of the boundaries of factors is user-dependent. Furthermore, the definition of the parameter space is particularly critical. This is usually done heuristically, suggesting

that non-ideal experimental settings are not necessarily identified and the parameter space has to be iteratively reduced step by step. Narrowing down the design space by using statistical DoE requires a lot of time and experimental effort, especially in cases where a high number of relevant factors are targeted. At the same time, the experiments can be limited in their information content, constraining the outcome of the optimization studies [8–10]. This generally results in a small increase in process knowledge only.

To reduce the number of experiments and increase the process understanding during the design and optimization of bioprocesses, a novel model-assisted design of experiments (mDoE) concept was recently introduced [11–13]. It combines the benefits of statistical DoE with a mathematical process model as a virtual representation of the bioprocess, called a digital twin. Although the term "digital twin" has not yet been defined across different parts of the industry, in bioprocesses they are intended to be a virtual counterpart of the bioprocess for the entire life cycle of the biopharmaceutical production process. In the context of mDoE, digital twins consist of a mathematical process model, which have gained increased importance in the last decades. They can be applied to design [14–16], control [17–19], and optimize [20, 21] biopharmaceutical production processes. The main intention of a mathematical model is to find solutions by analyzing the model in order to propose targeted experiments [22]. As they contribute to a scientific understanding of the process variables and their impact on the final product, mathematical process models in the field of biopharmaceutical production processes are now considered to be a sustainable part of QbD [7, 12, 23, 24].

2 Design of Experiments Methods

Even if traditional trial-and-error and one-factor-at-a-time methods are still used, advanced statistical DoE methods are applied more frequently in the field of biopharmaceutical process development [25–27]. They can be used for the statistical and systematic planning of experiments for hypothesis testing and/or the optimization of process variables (namely, "factors") with regard to the desired outcome, called "response" (e.g., product titer, product quality) [7, 28, 29]. In general, the process development based on DoE methods leads to a certain reduction in the number of experiments to be done in practice compared to one-factor-at-a-time approaches. In the context of designing biopharmaceutical production processes, they were used in the upstream as well as in the downstream part. As an example for the design of a bioprocess, Zhang et al. (2013) implemented a screening design to identify active parameters for the development of a serum-free medium for the cultivation of a recombinant CHO cell line. Afterward, the process parameters were optimized, and a fed-batch strategy was designed [30]. As an example for the part of product purification, Horvath et al. (2010) used a screening design with eight experiments to determine the effect of different process parameters on the isoelectric

point of a therapeutic antibody expressed in CHO cell culture. The pH, temperature, and the time of the temperature shift were significant. These factors were evaluated in three levels in a concluding response surface design to optimize the isoelectric point [31].

Statistical DoE methods are solely based on user-defined selections of the experimental design and the definition of factor limits, including the definition of experimental variables and their evaluated levels [8, 32, 33]. This can lead to error-prone decisions, iterative re-adjustments of the experimental space with several rounds of costly and time-intensive experiments, and even to a design that simply cannot be implemented [7]. Expert knowledge is required to select suitable boundary values for process development and optimization using DoE [7, 34–36]. Therefore, the combination of digital twins with DoE in mDoE offers a novel tool for the knowledge-driven development of bioprocesses.

2.1 Screening Designs

Screening designs are intended to identify the significantly influencing factors from a list of many potential factors [33, 37]. Therefore, different experimental designs can be used. The most commonly used designs, called full factorial, factorial fractional, as well as Plackett-Burman designs, are discussed.

2.1.1 Full Factorial Designs

A full factorial design can be used to examine the main effects and interactions of one or more factors on the respective response. The design consists of two or more factor levels and *k*-factors, resulting in at least a 2^k-design [38, 39]. Exemplary, the full factorial design for three factors is given by a 2^3-design plan, shown in Fig. 1a.

2.1.2 Reduced Full Factorial Designs

In order to reduce time-consuming and costly experiments in the case of a large number of factors, incomplete designs, like fractional factorial and Plackett-Burman designs, can be chosen. The fractional factorial designs, representing a reduced form of the two-level factorial design, are based on the assumption that higher-value interactions are irrelevant. This results in a 2^{k-n}-design, whereby the 2^k-design is reduced by n steps [38, 40, 41]. A reduced form of the previously mentioned 2^3-design plan, a fractional factorial 2^{3-1}-design, is shown in Fig. 1b. Plackett-Burman designs, a special form of the two-level fractional factorial designs, are suitable if the

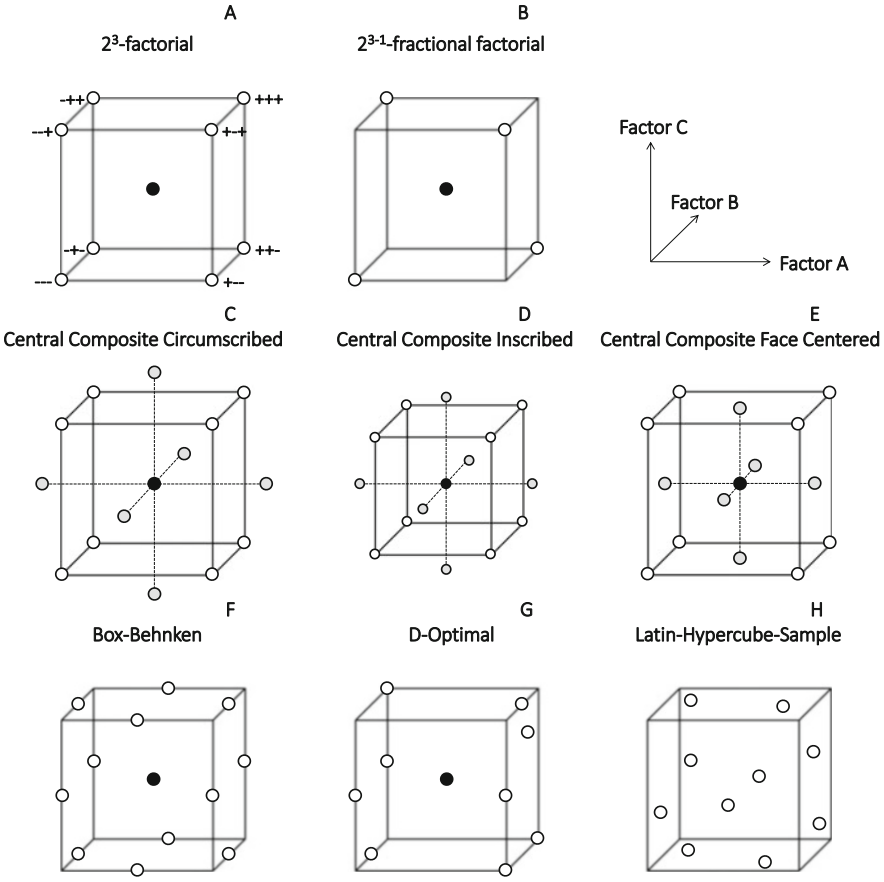


Fig. 1 Geometrical representation for screening (**a**, **b**) and optimization designs (**c**–**h**) and optimization designs with three factors (Factor **a**, Factor **b,** and Factor **c**). Dots represent the recommended experiments. The gray dots are the star points, and the black dots are the central points. All designs are examined at two levels (+ and -)

focus is on the investigation of the main effects and interactions can be disregarded [40]. However, a mixing of the effects can occur [38, 41].

2.2 Optimization Designs

In order to maximize a response, the levels of the influencing factors are optimized in so-called optimization designs. Therefore, the most known designs, like the central composite, Box-Behnken, optimal, and space-filling designs, are briefly introduced in the following.

2.2.1 Central Composite Designs

Central composite designs (CCDs) are built from factorial 2^k- or fractional factorial 2^{k-n}-designs. Additionally, center and star points (marked gray) are augmented, allowing the estimation of curvature. In general, three variations of CCD exist, which differ in range settings of their factors. Figure 1 C–E illustrates the relationships among these variants. Depending on the variant, the design is spherical, orthogonal, rotatable, or face centered [28, 38, 42].

2.2.2 Box-Behnken Designs

Box-Behnken designs (BBDs; see Fig. 1f) are based on the combination of a two-level factorial design with a balanced incomplete or partial block design [43]. They are nearly rotatable and require an examination on three levels for each factor, resulting in a field with distinct resolution of interactions and quadratic effects [44]. However, for a large number of factors, this implies a poor estimation of the two-factor interactions [43].

2.2.3 Optimal Designs

With optimal experimental designs, the experimental space can be restricted, and user-specific settings can be made. There are a large number of optimization criteria to distribute points in the experimental space. The most frequent representatives are average (A)-, determinant (D)- (shown in Fig. 1g), eigenvalue (E)-, global (G)-, and variance (I)-optimality. If the coefficients of the regression model are of interest, then A-, D-, and E-optimal plans are used. G- and I-optimality, however, refer to the fitted regression model [38].

2.2.4 Space-Filling Designs

Traditional experimental designs, such as the CCDs, BBDs, and the optimal experimental designs, often create experiments close to the factor boundaries. This can cause areas of free space, which are not examined and only minimize noise [45]. However, to minimize bias, space-filling designs can be used. In this case, possible experiments are randomly distributed in the individual spaces. An example of such designs is the Latin Hypercube Sample Design (LHSD), which fills the room evenly, allowing for a large number of factors and levels to be used (Fig. 1h). The experimental space is filled in such a way that there is an even distribution in the entire factor space or the maximum distance between the design points is minimized. However, the corners of the factor space are left out obtaining this information would only be possible by extrapolation [41, 46].

Table 1 Different designs for screening and optimization of CHO cultivation processes

Design	Opportunities	Challenges	Reference
Plackett-Burman	Development of a serum-free medium for the production of erythropoietin by suspension culture of recombinant Chinese hamster ovary cells	Confidence levels of 80% and obscure significant factors	[47]
Factorial	Identification of the demand for growth factors in the initial medium design, serum-free adaptation, stability analysis, and scale-up	No investigation of center points	[48]
Fractional factorial	Investigation of the effect of medium and feeding components on the main quality characteristics of a monoclonal antibody	Variation in statistical variance and different regression models	[49]
Optimal (D-optimal)	Development of a cultivation feeding protocol (feeding volume, starting point, time of shift in temperature, and osmolality)	High number of experiments to be performed experimentally	[34]
CCD	Optimization of the concentration and temporal addition of valproic acid (VPA) in three different CHO cell lines		[50]
BBD	Optimization of the amino acid combinations to determine the most effective concentration in the feed		[51]

2.3 Examples and Challenges of Conventional DoE

In this part, challenges of conventional DoE are discussed focusing on specific studies. A number of possible applications of screening and optimization designs are shown in Table 1.

Plackett-Burman designs are common designs for screening experiments. They are used, e.g., to identify the effects of amino acids and other components in conventional cell culture media formulations. Lee et al. [47] developed a serum-free medium for the production of erythropoietin by suspension culture of recombinant CHO cells, identifying six active determinants (glutamate, serine, methionine, phosphatidycholine, hydrocortisone, and pluronic F68) for cell growth. 79% of the erythropoietin titer achievable in the medium supplemented with 5% dialyzed fetal bovine serum were reached in the serum-free medium. However, 80% confidence levels were used to achieve useful statements, and some of the significant variables are obscure (e.g., pluronic F68) [47]. Chun et al. [48] used a full factorial design to identify effective growth factors in culture medium. Four growth factors were investigated on 2 levels, resulting in the implementation of 16 experiments. Important growth factors were identified. However, no center points were investigated; thus no curvatures could be detected [48]. Rouiller et al. [49] investigated six CHO cell lines in two different cultivation media to which six components were added in three different levels to develop a process for the production of monoclonal

antibodies. A two-stage fractionated factorial design with six factors was implemented, and various regression models were used to identify the active variables [49]. This resulted in 384 experiments to be performed, which was only possible by using a deep well plate system. Nevertheless, there were variations in statistical significance, and possible active variables have to be tested on a larger scale [49].

The amount of experiments to be performed can be seen as the main challenge in using optimization designs as well. The most commonly used optimization design is CCD. Yang et al. (2014) used a CCD to optimize the concentration and timing of valproic acid (VPA) addition to the cultivation of three different CHO cell lines [50]. Even the investigation of two factors for one cell line results in eight experiments. Torkashvand et al. (2015) optimized the concentrations of four amino acids (aspartic acid, glutamic acid, arginine, and glycine) in the feed using a BBD. The factors were investigated at 3 levels, resulting in 29 experiments to be implemented [51]. Duvar et al. (2013) developed a feeding protocol for a fed-batch CHO cultivation. The choice of a D-optimal experimental design resulted in 18 experiments with 4 factors (feeding volume, starting point, time of shift in temperature, and osmolality) [34].

For the previously mentioned studies, the planned experiments in statistical DoE result in identifying active parameters and optimization of the bioprocess. However, there are still challenges, and the implementation of statistical DoE can lead to time-consuming and costly rounds of experiments, especially if they are implemented in fed-batch mode. Furthermore, the heuristic selection of, e.g., the parameter settings or the design selection is seen critically. These rely on user-defined settings and mostly require a lot of time and experimental effort. But, as the investigation of various studies has shown, no adequate justification for the choice of an experimental design is provided. In addition, in conventional DoE only the experimental endpoints are examined, and therefore only the integral of it is judged. The entire time trajectory, with, e.g., metabolite formation or substrate uptake, is hardly reflected.

3 Model-Assisted Design of Experiments

The combination of statistical DoE with mathematical process models is a novel tool – enabling a knowledge-driven bioprocess development in the context of QbD. Using this method, the abovementioned limitations of DoE methods can be avoided, and the design as well as the optimization of bioprocesses can be improved. However, in contrast to the chemical industry, bioprocess design on the basis of mathematical models is not yet well established in biopharmaceutical manufacturing processes with mammalian cells [52]. According to experiences of the authors from discussions and projects, the use of model-based innovative methods for process development has so far failed due to different reasons:

- It is based on the lack of knowledge of the potential and limits of model-based/model-assisted methods and "bad experiences" with them (e.g., due to unrealistic expectations).
- There is a lack of method integration for a consistent development strategy, which can be adapted to existing work processes, a suspected high (modeling) effort and doubts about the transferability of methods and models to other processes.
- The qualification profile of the involved personnel does often not fit (required: biotechnology, process technology, modeling, and statistics).

An additional challenge is the application of models on complex metabolic pathways of mammalian cells regarding cell growth and product formation. In addition, models targeting the metabolism of cell cultures demand more effort than those applied in chemical or microbial processes. Even if mathematical models are a promising tool for the development of stable processes that comply with the principles of QbD, examples have so far only been published in the field of product purification and polishing [8]. Nevertheless, Möller et al. (2018) and Abt et al. (2018) showed that model-assisted and model-based DoE methods have great potential for the development of process strategies and makes the process development more knowledge-based [7].

General differences between model-based and model-assisted DoE methods are due to the aim of the recommended experiments. Model-based DoE (MBDoE) [53–55] is used to supply valid experimental data for a precise model structure and model parameter identification, where the conventional statistical DoE could fail [56]. Uncertainties are key information in MBDoE, as model and data imperfections cause undesirable variations in model parameters and simulation results. This variation drives the MBDoE methodology, where it manifests itself as optimal experimental settings (e.g., measurement principle, sampling rate, inputs/stimuli) and informative data [55, 57]. However, uncertainties cause a discrepancy between computed and experimental outputs leading to suboptimal or even meaningless experimental designs for model parameter adaptation. To overcome these problems, a sequential approach, as shown in [7], has proven to be very effective by increasing the robustness of the MBDoE against parametric uncertainties [58–60].

In model-assisted DoE, a process-related target (i.e., product titer) is optimized, and the model supports in the evaluation and recommendation of DoE designs. A structure for a model-assisted DoE concept is shown in Fig. 2. At first, a mathematical process model is used to describe, e.g., the growth, the substrate, and metabolite concentrations as well as the productivity of a specific cell line. Therefore, the model is adapted to first cultivation data (Fig. 2, Box 1), e.g., based on literature and/or existing knowledge. The evaluated data should be used to cover typical known effects, e.g., inhibitions or limitations. Certainly, the number of experiments that can be performed at this stage, preferably in small scale, such as shaking flasks or deep well plates [11–13], is usually limited. However, only a few experiments are required to generate the mathematical model, as shown in the case study (see Sect. 4). Accordingly, the number of experiments in mDoE is still less than the number of experiments to be performed in statistical DoE. Based on these data, model

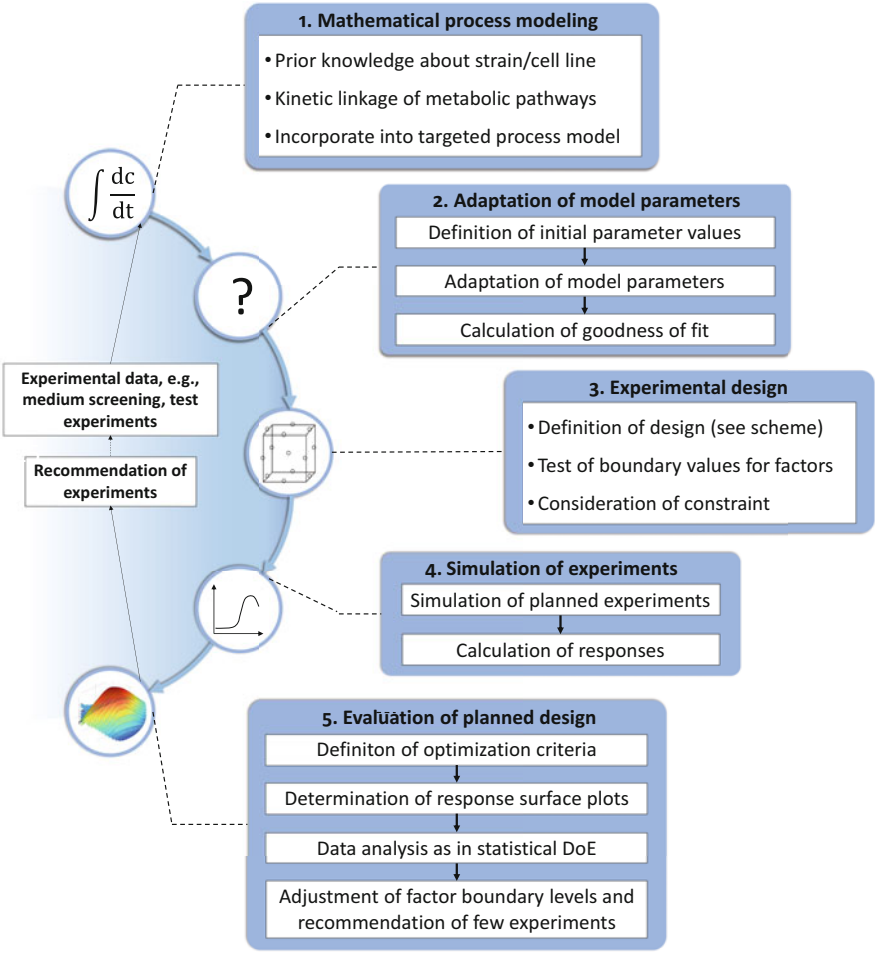


Fig. 2 Structure of the model-assisted design of experiments concept [11, 61]

parameters are adapted (Fig. 2, Box 2). Afterward, a statistical DoE design (see Sect. 2) is chosen (Fig. 2, Box 3). A scheme to select a design is explained in Sect. 3.2. The model is then used to simulate the responses for each previously planned experiment (Fig. 2, Box 4). Subsequently, the initial DoE is evaluated with respect to the defined factor boundaries as well as the experimental design (see Fig. 2, Box 5). This enables the testing of different designs, boundary conditions, optimization criteria, and factor as well as response combinations in silico before experiments are experimentally performed. This can be used to evaluate the mDoE method as well as boundary conditions and significantly reduce the number of experiments. Additionally, different designs can be chosen and computationally evaluated using the model simulations.

3.1 Digital Twins in Model-Assisted Design of Experiments

As already discussed, the term "digital twin" is still not sufficiently defined and has different meanings in different parts of industry. Historically, it is a computational model of a machine tool or a mechanical manufacturing site, and it is used to handle the increased complexity [8]. In the bioprocess industry, digital twins progressively include multiple parts of the manufacturing steps and their interaction [9]. They are intended to be a universal tool for the entire life cycle of a bioprocess, whereby the digital twins are virtual counterparts to the processes. They enable predictive manufacturing, meaning that bioprocesses can be analyzed, optimized, forecasted, and controlled [62]. The complexity of digital twins highly depends on the desired focus of application, and they can be based on a variety of complex structures as data-driven models, artificial neural network, or mathematical process models [22, 24].

With respect to the application of mDoE, the mathematical modeling in the initial phase of process development is, in the author's opinion, the starting point for knowledge integration into a digital twin for the entire life cycle of the bioprocess. The mathematical process model in the digital twin incorporates the process understanding, for which the degree of model complexity can be increased stepwise throughout the performed studies, as represented in Fig. 3. In this context, the

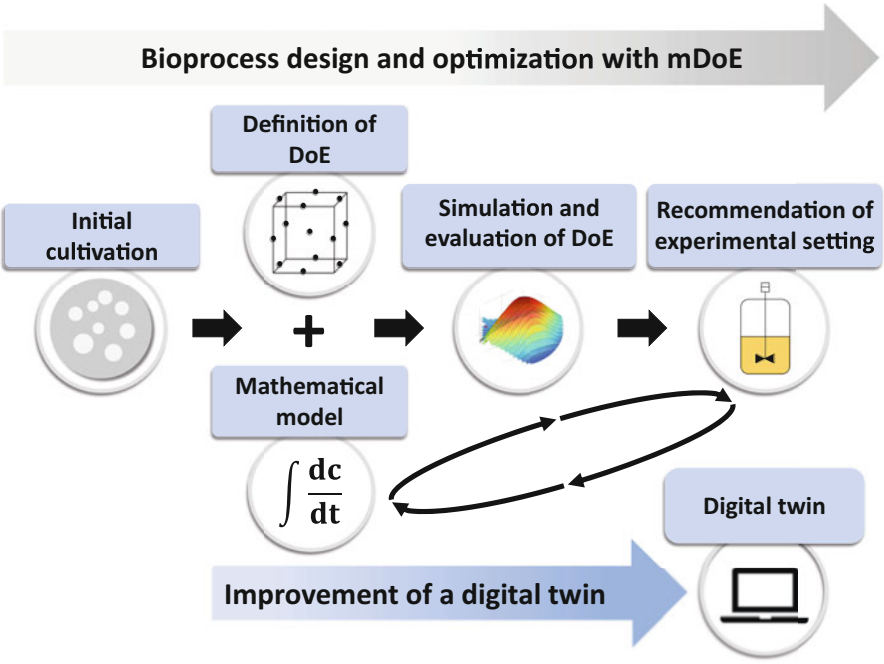


Fig. 3 Usage of the mDoE in bioprocess design and optimization as well as the development of a digital twin from a mathematical model

model structure should be kept as simple as possible in the initial process design phase and should then be extended if more data becomes available or novel biological effects are identified.

In the application of digital twins within mDoE (see Fig. 3) during process design and optimization with only a low number of available data, they structurally include rather simple mathematical process models. These are based on the formulation of mathematical links (i.e., equations) between cell growth, metabolism, and corresponding product formation [63]. If partial mechanistics are unknown, they can be modeled to gain a systematic understanding, although they might not be measurable (e.g., in systems biology) [17, 41]. Such a mathematical process model mainly works as an initial starting point to obtain a deeper process understanding during the bioprocess life cycle.

3.1.1 Mathematical Model Structures

Mathematical modeling has already been the subject of controversial discussions in recent years, and several models of varying complexity have been described in literature [64–66]. In the early phases of bioprocess development, the mathematical models used in mDoE mainly consist of simple model structures and should then be extended stepwise. The model parameters considered should be determinable by simple experiments since these include known mechanistics (e.g., ammonia formation based on glutamine uptake). It is favorable if models used for process optimization are applicable to a broad range of bioreactor scales [12, 64, 67]. Although the application of mathematical process models for the development of sophisticated processes has many advantages, it is still not commonly applied in bioprocess development. Reasons for this include the variety and complexity of mathematical models, e.g., different mechanistics and quality of predictions (recently reviewed in [64]). Due to the complexity of biological processes, simple models might be unsuitable for representing real phenomena. However, it has been suggested that the growth of a cell line follows the same kinetics regardless of the cultivation method, such as batch and fed-batch processes [65]. Nevertheless, even with complex models, the behavior of cells may change, and predictions can differ from observed behavior. Reasons are the inadequate precision of the approximated model coefficients and the complexity during the determination of the model parameters. Therefore, a compromise between the accuracy of the model and the required experimental effort for the determination of the parameters needs to be agreed on for each application [68].

Bioprocess-related mathematical models are either classified according to the description of the biophase, which is seen as an engineering-type approach or based on the implemented model structure (e.g., neural networks, fuzzy logic). This chapter focuses on biophase-classified models, which are historically sorted according to their structural complexity, as shown in Fig. 4. Even if this classification was made in the 1990s, it is still valid for the class of models here discussed.

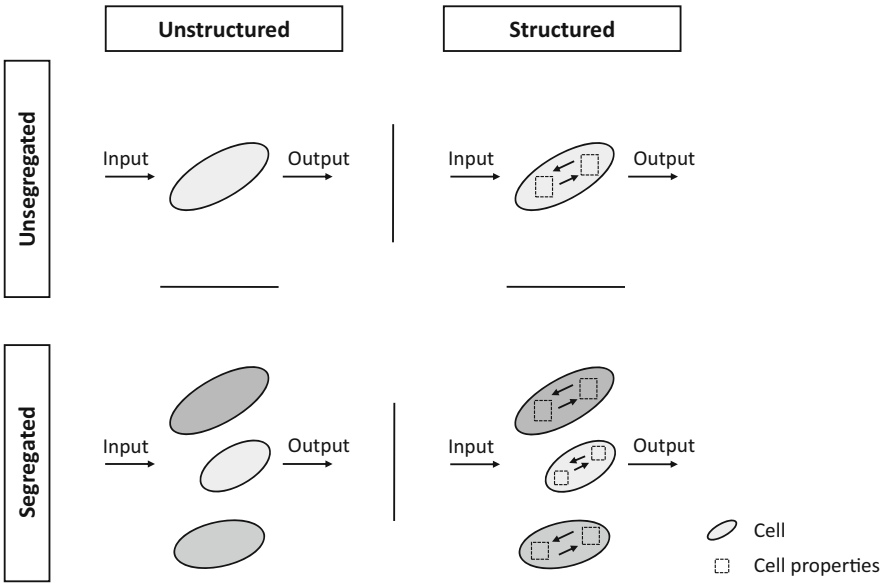


Fig. 4 Classification of mathematical models separated in unstructured, structured, unsegregated, and segregated

Unstructured and unsegregated models describe the biophase as one component and use kinetic equations to describe their interaction and response to the environment, e.g., the effect of glucose concentration on bulk cell growth. They are widely applied for industrial applications and are state of the art [69]. It is advantageous that the model parameter estimation is based on only a few measured concentrations [70]. Moreover, a method for their knowledge-driven development was recently reported by Kroll et al. (2017) [23]. With the development of novel analytical methods, *structured and unsegregated* models were developed. The cellular properties are reflected by average cells with the same physiological, morphological, and genetic identity [71–73]. They aim to describe intracellular metabolic pools in otherwise average cells. Most examples try to examine the intrinsic complexity of cell metabolism. Lei et al. (2001) described the growth of *Saccharomyces cerevisiae* based on glucose and ethanol using two modeled pools, which describes the catabolism and anabolism, respectively [74]. Moreover, a six-compartment model for microbial and mammalian cell culture was recently introduced to reduce the modeling effort as a basis for digital twins [75]. Flux balance analysis, mostly used in systems biology, is additionally associated with the structured model class [76, 77].

In *unstructured and segregated* models, different separated cell populations are modeled with the description of the metabolism by bulk kinetic equations [78]. The scope of application is broader, leading to the determination of cell culture quality and gaining an understanding of the cell cultivation process. Exemplary, cell-cycle-dependent population balance models were introduced [22, 24, 78, 79]. Therefore,

different cell-cycle-dependent growth rates, metabolic activity, and DNA replication rates are modeled, and metabolic regulations were studied, but the degree of complexity and computational power increases significantly from a few seconds to multiple hours. Hence, they require a comprehensive knowledge of the mechanisms and more data to estimate the model parameters. *Segregated and structured* models describe the nature of cell cultures with individual single-cell metabolism and their interaction with the medium. Sanderson et al. (1999) introduced a single-cell model that describes the interaction of 50 components in the medium, cytoplasm, and mitochondria for an antibody-producing CHO cell line [80]. Other examples could be found for baculovirus-infected insect cell cultures [81] and the amino acid metabolism of HEK293 and CHO cells [82]. However, the computational power and amount of data required to estimate the model parameters are still considerable, which can limit their industrial application.

3.2 Recommendations on the Selection of Designs for mDoE

The choice of an experimental design significantly influences the implementation of DoE and mDoE. Usually, the selection of a design depends on basic settings (number of factors, number of factor steps, the regression model, and the number of test runs) and design-specific properties (block formation, orthogonality, and rotatability). However, as the investigation of various studies has shown, in most references less information for the choice of an experimental design is provided. DoEs are mostly selected based on heuristics within a given scientific field, and there is no guided decision-making workflow yet. Based on the author's understanding, a scheme (Fig. 5) is presented in the following to assist in the selection of appropriate DoE designs in the field of bioprocess engineering. This scheme was developed based on literature and is seen to assist in the selection of DoE designs within mDoE [8, 11, 12].

Due to their favorable properties, CCDs and BBDs are most frequently used for optimization [39]. If settings are adjusted individually, optimal designs should be used. As a result of the low computational effort, the D-optimal design has become generally accepted among the optimal designs [38]. Therefore, commercial software tools for creating optimal designs are often limited to the D-optimal design [83]. However, the I-optimal designs are sometimes recommended [39, 83]. If a large area of the factor space is to be covered, it is recommended, e.g., to combine the LHSD with an optimal design. This results in a better distribution of points across the factor space and reduces both bias as well as noise [45].

In the first decision-making level, the number of investigated factors k is used. Except for BBDs, which require at least three factors, the number of factors can be selected as desired [38, 41, 84]. Typically, three to six factors are used for processes optimization [38]. In the case of a high number of factors and the use of insignificant factors, the LHSD is recommended. This is enabled by the random distribution of experiments. Hence, if one or more factors appear not to be important, every point in

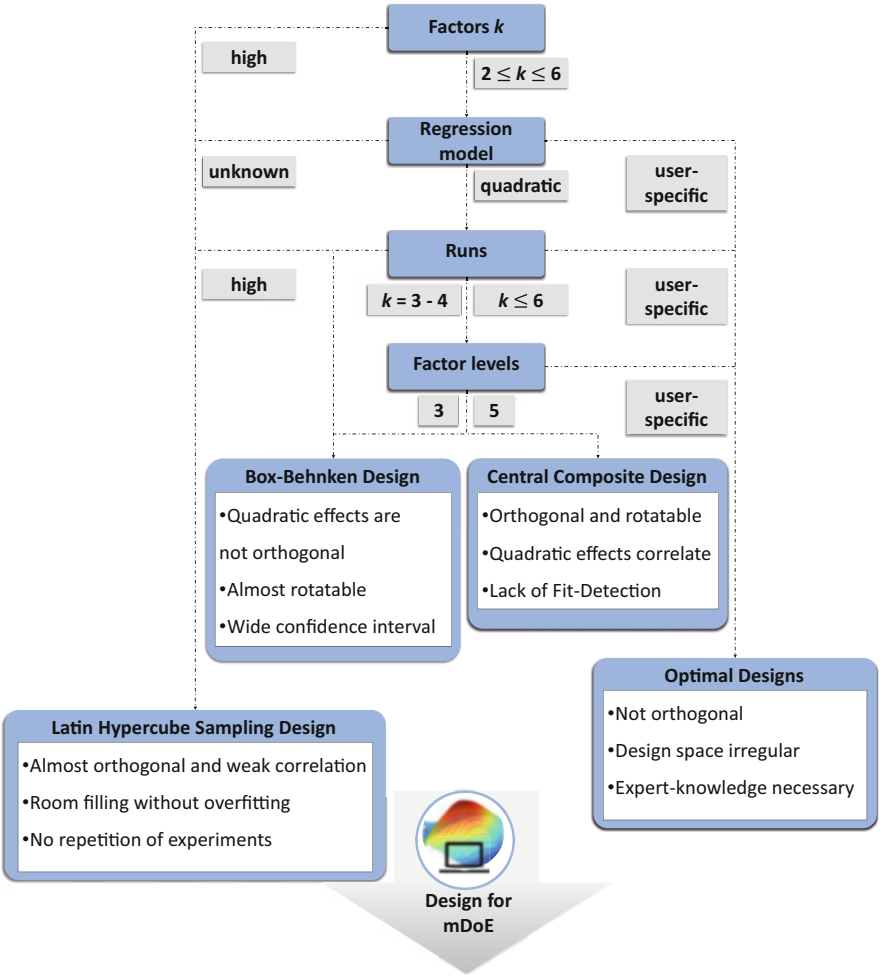


Fig. 5 Scheme for selection of designs in mDoE

the design still provides some information regarding the influence of the other factors on the response [41]. However, in bioprocesses the number of factors is significantly larger than the observations. Therefore, domain knowledge as used in mDoE is needed, captured on models as additional constraints to the system.

On the next level, the desired regression model is defined. Generally, the use of a quadratic regression model is recommended, since higher-order regression models lead to an increasing number of unknown coefficients and can lead to an overfit [38, 41]. For CCDs und BBDs, a quadratic regression model is used by default, whereas for optimal designs, the regression models can be user-defined. If the regression model cannot be defined, the LHSD can be used [39].

The number of runs is taken into account at the penultimate level. BBD is most efficient with three or four factors. Compared to other designs, they require the least number of runs [44]. However, even with optimal designs, the number of experiments can be user-defined and thus minimized. Only a model-dependent minimum number must be included. For a quadratic regression model with four factors, e.g., the minimum number is 15, one term for the intercept, four linear terms, four purely quadratic terms, and six cross-product terms must be taken into account. In the BBD or CCD, the number of runs varies between 25 and 30, depending on the number of center points [85]. LHSDs, on the other hand, require a large number of experiments to fill the space and are therefore mainly used for computer simulations [86–88].

Finally, the choice of factor levels is decisive. In order to evaluate designs optimally, they require at least three levels per factor [43]. In the BBD, the factors are examined at three levels. However, no factor-level combinations are investigated at the corner points, and thus, only a low prediction quality for extrema is available [89]. LHSDs are also not suitable for the investigation of extrema [41]. The CCDs contain five levels per factor [28]. In the optimal designs, the factor levels can also be set user-specifically and combined as desired. However, it can happen that very few test points are generated in the middle of the test area, which means that no statements can be made concerning this area [38].

The selected design can then be implemented experimentally or in the mDoE (see Sect. 4.3). It should be noted that the application of the selection scheme (Fig. 5) is not limited to mDoE solely and it can be generally applied for the selection of DoE designs.

4 Case Study: mDoE for Medium Optimization

As previously mentioned in Sect. 3.1, digital twins are used within the mDoE concept to simulate and evaluate statistical DoE designs in silico. The application of mDoE with a strong reduction in the number of experiments has been shown so far for medium optimization, fed-batch design, and scale-up studies for antibody-producing CHO cells, algae, and yeasts [11, 12, 61, 90]. During these studies, the process understanding is stepwise increased and captured in the digital twin (i.e., mathematical process model). In the following, the application of mDoE is exemplarily discussed for the reduction of the factor boundary values for the optimization of the glucose and glutamine concentrations in an antibody-producing cell culture process. The specific workflow applied in this study is shown in Fig. 6.

In this case study, the dynamics of the bioprocess are modeled first, and the model parameters are based on a few experimental data points. Then, the boundary values of experimental designs are defined, and experimental settings are planned. Each planned experiment is simulated, the responses (e.g., maximal product titer) are calculated, and the response surfaces are determined. Based on these response surfaces, the initially defined factor boundary values and the planned experiments are evaluated, and only a few experiments are recommended to be performed. This

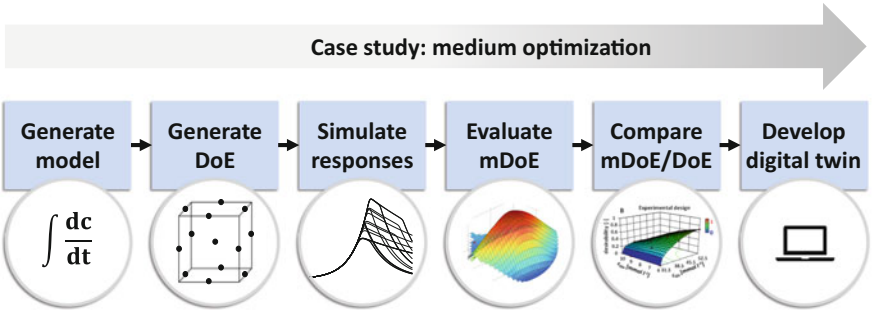


Fig. 6 Workflow of the upcoming chapters of the medium optimization in the case study

results in a significant reduction in the number of experiments to be performed. This example shown in the following is based on our previous publication, and more details can also be found in Möller et al. (2019) [11]. This medium optimization is only a small part of a process development workflow, which could be implemented from medium optimization over fed-batch design to scale-up using mDoE [12]. This resulted in the evolution of the digital twin, as briefly explained in Sect. 4.6.

4.1 Mathematical Process Model

In this case study, an unstructured, non-segregated saturation-type model was used as virtual representation of the bioprocess. The mathematical model from literature [19] was adapted and modified to describe the dynamics of cell growth and metabolism of antibody-producing CHO DP-12 cells in batch mode (see Table 2). This model was chosen due to its simple model structure and the opportunity to estimate all the model parameters from just a few shaking flask cultivations.

4.1.1 Batch Process Model as Digital Twin

According to the mDoE workflow (Fig. 2, Box 1), the mathematical process model is used to simulate the growth of the CHO DP-12 cells. It is based on the linkage of the main substrates glucose (c_{Glc}) and glutamine (c_{Gln}) as well as the main metabolites lactate (c_{Lac}) and ammonium (c_{Amm}) to describe the behavior of the cells (X_t, total cell density, and X_v, viable cell density). Cell growth is modeled with kinetic parameters $K_{S,i}$ (i = Glc, Gln), a maximal growth rate (μ_{max}), a cell lysis constant (K_{Lys}) of dead cells, and a minimal ($\mu_{d,\,min}$) and a maximal death rate ($\mu_{d,\,max}$). Since no inhibition of cell growth could be detected in batch mode, inhibitory components were not considered in the model. Therefore, the calculation of the specific growth rate μ (Eq. 9, in Table 2) and specific death rate μ_d (Eq. 10, in Table 2) is based on a Monod-like structure of the substrates glucose and glutamine, with only the substrate

Table 2 Mathematical process model in batch mode, modified from [19]

Balance equations	Kinetic links
Biomass	
$\frac{dX_v}{dt} = (\mu - \mu_d) \cdot X_v$ (1)	$\mu = \mu_{max} \cdot \frac{c_{Glc}}{c_{Glc}+K_{S,Glc}} \cdot \frac{c_{Gln}}{c_{Gln}+K_{S,Gln}}$ (9)
$\frac{dX_t}{dt} = \mu \cdot X_v - K_{Lys} \cdot (X_t - X_v)$ (2)	$\mu_d = \mu_{d,min} + \mu_{d,max} \cdot \frac{K_{S,Glc}}{K_{S,Glc}+c_{Glc}} \cdot \frac{K_{S,Gln}}{K_{S,Gln}+c_{Gln}}$ (10)
$\frac{dVi}{dt} = \frac{\frac{dX_v}{dt} \cdot X_t - X_V \cdot \frac{dX_v}{dt}}{X_t^2}$ (3)	
Substrates and metabolites	
$\frac{dc_{Glc}}{dt} = -q_{Glc} \cdot X_v$ (4)	$q_{Glc} = q_{Glc,max} \cdot \frac{c_{Glc}}{c_{Glc}+k_{Glc}} \cdot \left(\frac{\mu}{\mu+\mu_{max}} + 0.5\right)$ (11)
$\frac{dc_{Gln}}{dt} = -q_{Gln} \cdot X_v$ (5)	$q_{Gln} = q_{Gln,max} \cdot \frac{c_{Gln}}{c_{Gln}+k_{Gln}}$ (12)
$\frac{dc_{Lac}}{dt} = q_{Lac} \cdot X_v$ (6)	$q_{Lac} = Y_{Lac,Glc} \cdot \frac{c_{Glc}}{c_{Lac}} \cdot q_{Glc} - q_{Lac,uptake}$ (13) $c_{Glc} < 0.5\ mmol^{L-1} : q_{Lac,uptake} = q_{Lac,uptake,max}$ (14)
$\frac{dc_{Amm}}{dt} = q_{Amm} \cdot X_v$ (7)	$q_{Amm} = Y_{Amm,Gln} \cdot q_{Gln}$ (15)
Antibody	
$\frac{dc_{mAb}}{dt} = q_{mAb} \cdot X_v$ (8)	$q_{mAb} = \gamma$ (16) $c_{Glc} < 1\ mmol\ L^{-1} : \frac{dc_{mAb}}{dt} = 0$ (17)

with the lowest concentration being relevant for growth. The cell-specific uptake rates of glucose and glutamine depend, in contrast to the growth, only on the current glucose and glutamine concentration (Eqs. 4, 5, 11, 12, in Table 2). However, the uptake rate of glucose is reduced at low concentrations. The concentrations of lactate and ammonium are proportional to the uptake rates of glucose (lactate) or glutamine (ammonium) (Eqs. 6, 7, 13, 15, in Table 2) and are linked with the yield coefficients ($Y_{Amm/Gln}$ and $Y_{Lac/Glc}$). In case of glucose concentrations below 0.5 mmol L^{-1}, a shift of lactate production to lactate uptake was considered (Eq. 14, in Table 2). The antibody production (Eqs. 8, 16, in Table 2), according to Frahm et al. [19], describes the production proportional to the viable cell density. However, glucose concentrations below 1 mmol L^{-1} stop the antibody production (Eq. 17, in Table 2).

4.1.2 Adaption of Model Parameters

The initial experiments for modeling were based on the previous publications of Beckmann et al. [91] and Wippermann et al. [92] with the same medium and cell line. Biological experiments were performed in quadruplicates, the data were averaged, and the model was adapted as well as model parameters estimated (Fig. 2, Box 2). This initial adaption and the further use of the mathematical model in mDoE can be seen as the starting point into a digital twin. Therefore, the model parameters were adapted. To compare and evaluate the quality of adaption, the modeled simulations and cultivation data were plotted and the coefficient of determination calculated. If the values tend to 1, the behavior of the cells could be represented with high accuracy. However, the area, which should be optimally displayed, should be

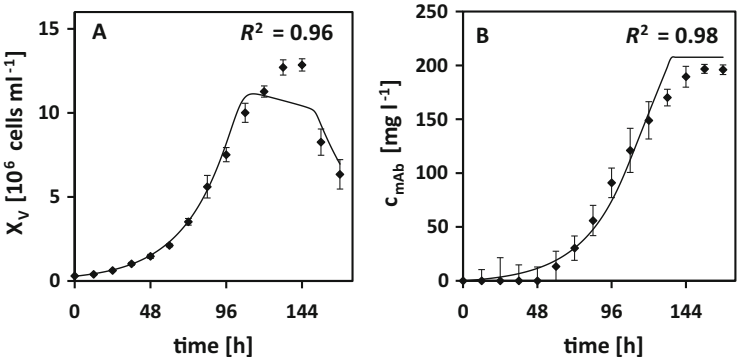


Fig. 7 Comparison of experimental data (◊) and simulated data (−), exemplary for viable cell (**a**) density and antibody (**b**). Mean and one standard deviation of four parallel batch cultivations. The samples were measured as three technical replicates for each shaking flask, and R^2 was calculated compared to the mean experimental data points

focused on. The goal could be a high cell density; therefore the representation of the stationary phase and the death phase is not important. This must be taken into account during the evaluation. Alternative measures are detailed explained in literature, e.g., [93, 94].

The simulation with the adapted model parameters is exemplary shown for cell growth and antibody production in Fig. 7. The exponential cell growth, the transition to the stationary phase, and the death phase could be simulated with an accuracy of $R^2 = 0.96$ (Fig. 7a). The antibody concentration increases until X_v decreases after approx. $t = 144$ h and was estimated with a high accuracy of $R^2 = 0.98$ (Fig. 7b). By this, the previous knowledge is captured into the model structures, and the model parameters reflect the cell behavior, which could further be used in mDoE.

4.2 Selection of Experimental Design

The determination of a suitable design is essential for the most appropriate evaluation of the mDoE and thus the optimization of the respective process, as can be seen in Fig. 2, Box 3. Therefore, a design for the mDoE was chosen considering the scheme in Fig. 5. In the first decision-making level of the scheme, the number of investigated factors k, which are two in this case study, was examined. Since BBD requires the use of at least three factors and LHSD is recommended for a high number of factors, only the CCD and optimal experimental designs remain. Then, the regression model was considered. For both, CCD and optimal designs, the recommended quadratic regression model can be used, although this is not adjustable for CCD. Therefore, no further restriction has yet been possible on the basis of this level. Finally, the third level can be used to select the design. At this level the number of runs is taken into account. Since the number of runs should be set

individually, the CCD was discarded, and an optimal experimental design was chosen.

Experiments were designed with suitable DoE software (in this study, Design-Expert 9, Statcon, USA), and each experimental factor combination of the experimental design was simulated (MATLAB). In the simulated design (20 experiments, D-optimal design), the initial glucose concentration was varied between 20 and 60 mmol L^{-1}. This corresponds to a 50% increase/decrease of the glucose concentrations related to the standard medium formulation as reported in Beckmann et al. [91] and Wippmann et al. [92]. These studies did not focus on the optimization of the batch-medium composition as aimed in this study. Glutamine concentrations typically applied in batch media range from 2 mmol L^{-1} up to 8 mmol L^{-1}. The factor range of the initial glutamine concentration was, therefore, widely defined between 2 and 12 mmol L^{-1}.

4.3 Simulation of Experiments

Each planned factor combination (Fig. 8a-d) and the corresponding responses (Fig. 8b, c, e, and f) were simulated using the mathematical model (as described in Fig. 2, Box 4). As can be seen in Fig. 8, the concentration range of the substrates

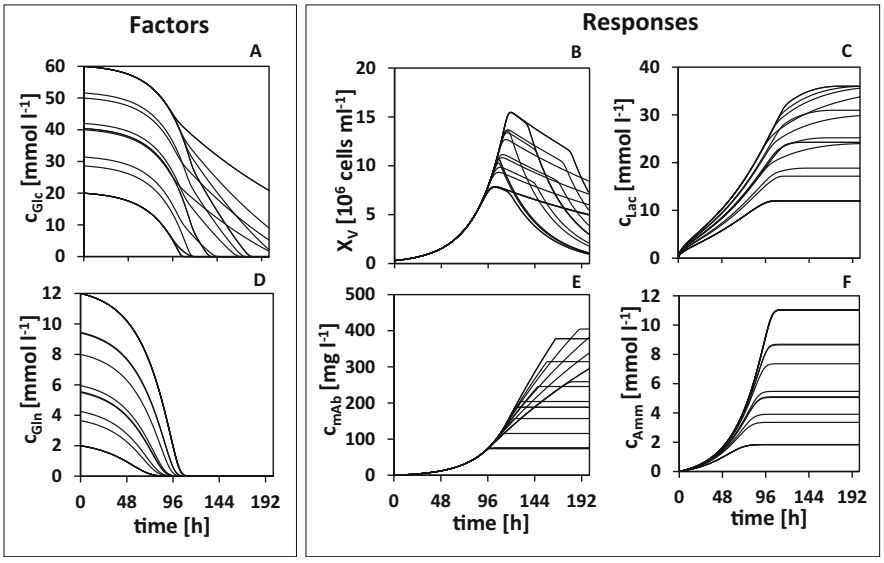


Fig. 8 Simulation of the growth behavior of CHO DP-12 cells for all initially planned experiments (Fig. 2, Box 3); (**a**) and (**d**), dynamic changes of investigated factors; (**b**), (**c**), (**e**), (**f**), investigated responses

(Fig. 8a, d) causes a maximum cell number between 8×10^6 cells mL^{-1} and 15×10^6 cells mL^{-1} (Fig. 8b). The maximum antibody concentration lies in a range between 75 mg L^{-1} and 380 mg L^{-1} (Fig. 8e). The simulated concentrations for lactate and ammonium range between 12 mmol L^{-1} to 35 mmol L^{-1} and 2 mmol L^{-1} to 11 mmol L^{-1} (Fig. 8c, f). By simulation, the time course is also determined, which is not the practice in typical experimental DoEs.

4.4 Evaluation of Planned Design

Out of the simulations, the maximal simulated values of X_v, c_{Lac}, c_{mAb}, c_{Amm} were exported as responses to generate response surface plots (Design-Expert 9). The simulated responses were treated in the same way as data from experiments. For this purpose, no data transformation was applied, and after analysis of variance (ANOVA, all hierarchical design mode, quadratic process order), an internal RSM was set up with a maximal significance value of 0.05. As can be seen in Fig. 9, different shapes of the response surface plots were determined, and their individual optimum is different, e.g., X_v is maximal for high initial glucose and glutamine, while c_{mAb} is high, regardless of the glutamine concentration. Such interaction could hardly be predicted before and was only possible through the simulations. After defining the RSM for each response, user-defined constraints for medium optimization were chosen (displayed in Fig. 9). The constraints were chosen to maximize X_v above a minimal X_v of 10^7 cells mL^{-1}. Furthermore, c_{mAb} should be maximized. The constraints for the metabolic waste products were defined based on the literature data with respect to cell growth and product quality. High lactate concentrations were shown to correlate with a reduced integral of viable cell density and a reduced product titer at day 14 in pH-controlled shaking flask cultivation with added sodium lactate [95]. Lactate concentration below 20 mmol L^{-1} is considered to not harm cell growth and productivity, whereby lactate concentration higher than 40 mmol L^{-1} was shown to harm CHO cell growth [96]. Therefore, a maximal c_{Lac} of 30 mmol L^{-1} was defined as the upper constraint, and the lactate concentration was minimized below this value to avoid potential lactate inhibition. The ammonium concentration was defined to be minimized. This was motivated based on the following understanding of its impact on product quality, even if it was not measured. Andersen et al. (1995) identified that the sialylation of a granulocyte colony-stimulating factor was significantly reduced by ammonium concentrations over 2 mmol L^{-1} [97]. Ha et al. (2015) investigated the mRNA expression levels of 52 *N*-glycosylation-related genes in recombinant CHO cells producing an Fc-fusion protein and observed a decrease of the protein production and the viable cell density after an addition of 10 mmol L^{-1} ammonium chloride. Simultaneously, the sialic acid content and the acidic isoforms were reduced after 5 days of cultivation [98].

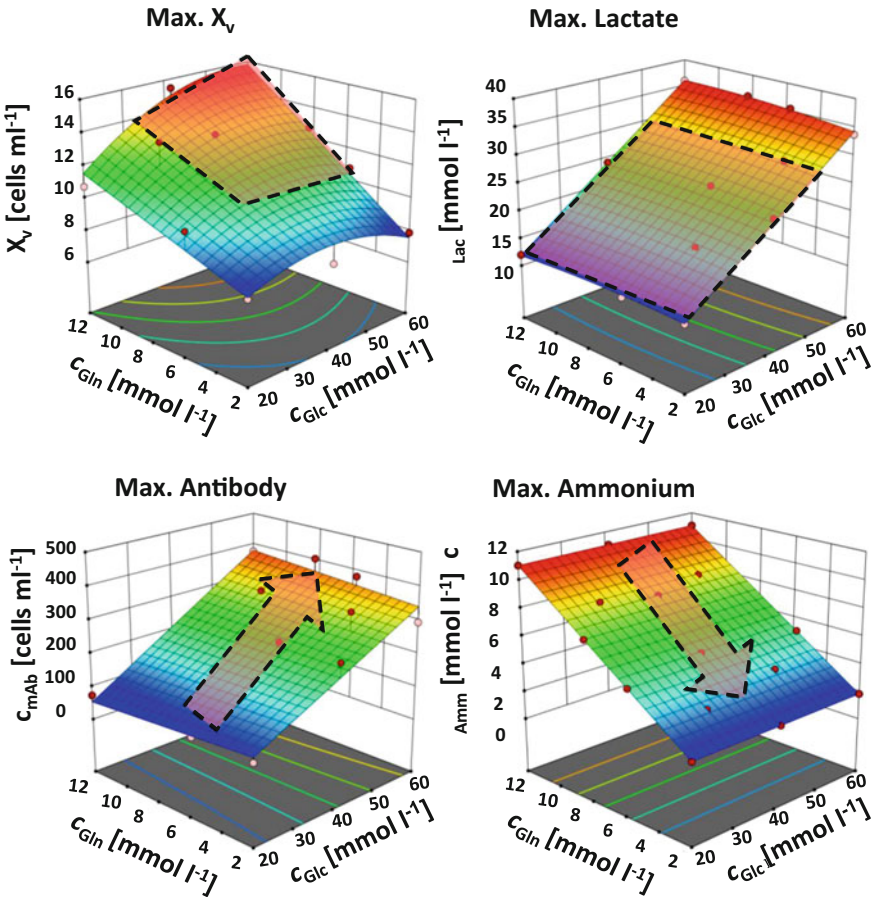


Fig. 9 Different shapes of the response surface plots for X_v, c_{Lac}, c_{mAb}, and c_{Amm} with different individual optimum. The user-defined constraints for medium optimization were marked red

Subsequently, an objective (i.e., desirability) function d_i was calculated for each response y_i individually, based on the user-defined constraints as lower acceptable response L_i and the upper acceptable response U_i.

$$d_i(y_i) = \begin{bmatrix} 0 & \text{if } y_i < L_i \\ \dfrac{(y_i - L_i)}{(U_i - L_i)} & \text{if } L_i < y_i < U_i \\ 1 & \text{if } y_i > U_i \end{bmatrix} \tag{18}$$

$d_i(y_i)$ is 0 if the optimization criteria is not fulfilled, and $d_i(y_i)$ tend toward 1 if the optimization is highly desirable. The multidimensional optimization problem is

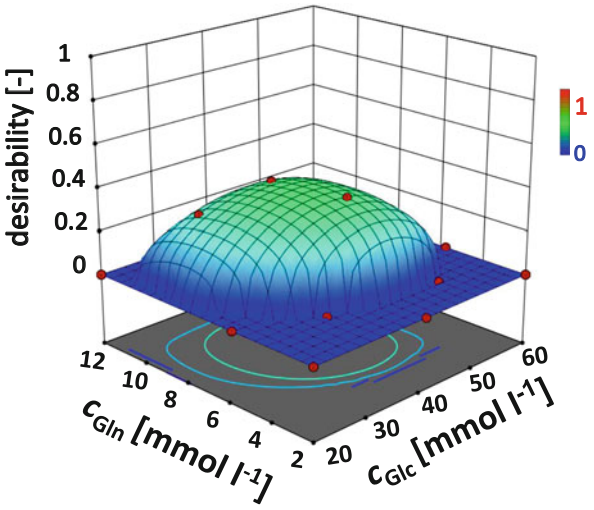


Fig. 10 Response surface plot for the desirability function of the factors glucose and glutamine. The optimization criteria are fulfilled if the desirability tends toward 1

reduced with the multiplication of the different desirability function values $d_i(y_i)$ to one overall desirability D:

$$D = \prod_{i=1}^{n} d_i(y_i) \tag{19}$$

The overall desirability function was calculated for the constraints mentioned and is shown in Fig. 10.

Glutamine concentrations higher than approximately 10.5 mmol L^{-1} and glucose concentrations above 52 mmol L^{-1} result in a optimization criteria $D = 0$. The optimization criteria were also not reached below 4 mmol L^{-1} glutamine and 21 mmol L^{-1} glucose. The performance of these experiments would be time- and cost-intensive, without providing sufficient knowledge. In this way, multiple constraints were considered and only a small area (5 of the 20 evaluated factor combinations) results as suggested experimental space with $D > 0$. Only this 5 factor combinations of the 20 evaluated would increase the process understanding.

4.5 Comparison to Experimentally Performed Design

The usage of mDoE allows the a priori evaluation and reduction of the boundary values if mechanistic links could be formulated beforehand. The reduced experimental space was selected within the estimated desirability function (Fig. 10). Based on the evaluation of Fig. 10, the boundary values for the initial glucose concentration were defined between 52.5 mmol $L^{-1} \geq c_{Glc} \geq 32.5$ mmol L^{-1}. The initial glutamine concentration has to be between 10 mmol $L^{-1} \geq c_{Gln} \geq 6$ mmol L^{-1}.

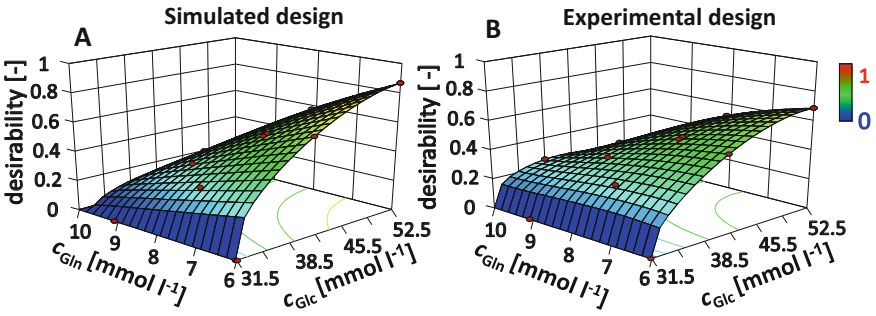


Fig. 11 Reduced simulated (**a**) and experimentally (**b**) performed DoEs. Points are the considered factor combinations

After the evaluation of the boundary values of the initially planned design, a new D-optimal design was planned (see Fig. 5) within the reduced design. Therefore, 16 experiments in a D-optimal design based on the reduced boundaries were planned and experimentally performed as well as simulated. The experimentally performed design was realized in 16 parallel shaking flask cultivations (approximately ten samples per cultivation). For the evaluation of the DoE in mDoE, the responses were only simulated. Both designs (experimental performed and simulated) were statistically evaluated, and the response surfaces were estimated. Both desirability functions were calculated due to the maximization of the antibody concentration and the minimization of the ammonium concentration and are shown in Fig. 11.

The evaluation is performed by the executing person, e.g., rely on their individual experience or user-defined constraints (device settings, etc.). Optimal starting concentrations in the upper right corner (high glucose as well as low glutamine concentrations) were recommended with $D = 0.87$ for the simulated design (Fig. 11a) and nearly the same for the experimentally performed design ($D = 0.70$). These small differences are typical when comparing the simulated results with uncertainty-based experimental results. No further experiments needs to be performed outside of this area, since the outer experimental space was evaluated beforehand using the digital twin (Sect. 4.4). Compared with the full experimental performed design, mDoE results in a reduction of 75% in the number of experiments (4 experiments for modeling vs. 16 experiments in experimental DoE).

The combination of model-assisted simulations with statistical tools can be used to decrease the experimental effort during medium optimization studies. Furthermore, the modeling study itself leads to an increase of the process understanding, which is part of QbD. No heuristic restrictions with several iterative rounds were necessary, because the mathematical process model incorporates the known factors and interactions and their dynamics in DoE. Furthermore, DoEs are typically based only on endpoints, and different responses and endpoints can be tested using the kinetic model.

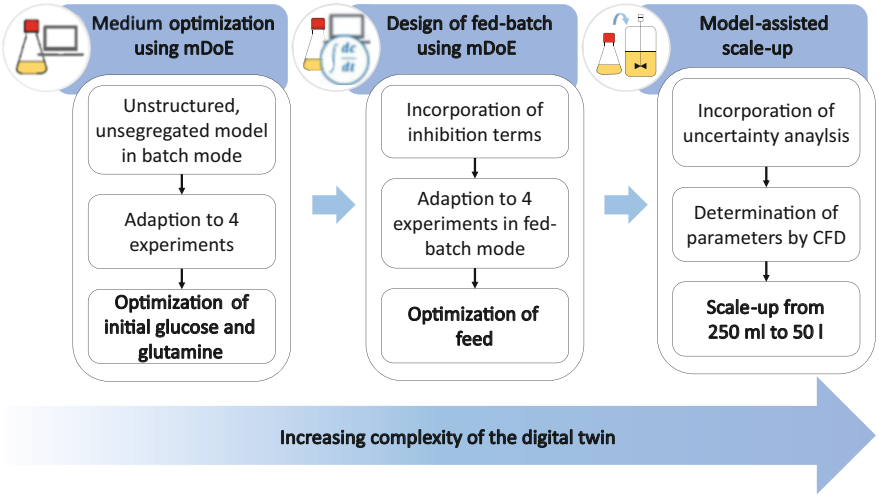


Fig. 12 Process development workflow in context to the digital twin

4.6 Further Development of the Digital Twin in Process Development Workflow

The evolution of the mathematical process model as digital twin is part of mDoE, as briefly focused in the following. As shown in Fig. 12, an unstructured, unsegregated model was initially adapted and modified to describe the dynamics of cell growth and metabolism of antibody-producing CHO DP-12 cells for the purpose of medium optimization in batch mode (Sect. 4.1). The model incorporated known mechanistic links for CHO cells, and the initial data for modeling was based on just four experiments, and optimal conditions for the medium composition were identified [11].

The complexity of the digital twin could be further increased during the process development workflow. The mathematical model was expanded by metabolic inhibition terms to optimize cell growth and productivity in fed-batch mode [11]. Therefore, the digital twin was used to optimize the concentration of glucose as well as glutamine in the feed, the feeding rate, and the start of feeding. After optimizing the medium composition and fed-batch strategy, the digital twin was used in model-assisted scale-up to evaluate the bioprocess dynamics during process transfer and scale-up computationally [12]. Therefore, the mathematical model was extended by model parameter probability distributions, which were determined at different bioreactor scales due to measurement uncertainty. Finally, the quantified parameter distributions were statistically compared to evaluate if the process dynamics have been changed and the former optimized fed-batch strategy was successfully scaled up to 50 L pilot scale. The application in these different processes has deepened the knowledge and thus steadily increased the complexity of the digital twin [11, 12].

5 Conclusion and Outlook

In this chapter, the mDoE concept for the combination of mathematical process models with DoE was described. The most commonly used designs were examined, and one representative study was described in detail. The role of a digital twin was discussed and applied to a medium optimization case study using mDoE. A mathematical process model as a starting point for digital twin was adapted to four experiments, and widely distributed boundary values for a DoE were evaluated using model predictions instead of laboratory experiments. The reduced experimental spaces were experimentally performed (DoE) and compared to the simulated DoE (mDoE). The same optimal conditions were found, and the further development in different steps of a process development workflow was described. Finally, the development of a digital twin and its use in mDoE can be seen as a useful tool in decision-making for process development and optimization with DoE in QbD.

Statistical DoE can still be used for initial screening studies and can also lead to process optimization in several rounds. Compared to conventional DoE, mDoE supplies a more knowledge-based development of bioprocesses. Due to the mathematical model in mDoE, challenges in DoE can be avoided. The mathematical model can be used for simulating the entire time trajectory with, e.g., metabolite formation uptake. Hence, not only endpoints of experiments are examined. Thus the knowledge about the process can be increased. Furthermore, domain knowledge is required and can be captured as additional constraints to the system, leading to a focused screening or optimization of bioprocesses using the mathematical model as a digital twin in mDoE.

Currently, the mDoE approach is tested for algae, yeasts, and cell culture. Further applications of digital twins and mDoE can be seen in the field of cell therapeutics, e.g., in the treatment of previously untreatable diseases as tumor diseases, brain insult, and chronic infections. Since, the production of cells is still mainly performed in static culture systems (e.g., T-flasks), it is difficult to provide a sufficient quantity of patient-specific cells. A digital twin in combination with mDoE could be used to build up an understanding of the process and, e.g., scale-up to enable fast and efficient proliferation of stem and immune cells.

References

1. Nelson AL, Dhimolea E, Reichert JM (2010) Development trends for human monoclonal antibody therapeutics. Nat Rev Drug Discov 9:767–774
2. Walsh G (2014) Biopharmaceutical benchmarks 2014. Nat Biotechnol 32:992–1000
3. Kretzmer G (2002) Industrial processes with animal cells. Appl Microbiol Biotechnol 59:135–142
4. Walsh G (2018) Biopharmaceutical benchmarks 2018. Nat Biotechnol 36:1136–1145
5. Chen C, Le H, Goudar CT (2016) Integration of systems biology in cell line and process development for biopharmaceutical manufacturing. Biochem Eng J 107:11–17

6. DiMasi JA, Grabowski HG, Hansen RW (2016) Innovation in the pharmaceutical industry: new estimates of R&D costs. J Health Econ 47:20–33
7. Abt V, Barz T, Cruz-Bournazou MN, Herwig C, Kroll P, Möller J, Pörtner R, Schenkendorf R (2018) Model-based tools for optimal experiments in bioprocess engineering. Curr Opin Chem Eng 22:244–252
8. Möller J, Pörtner R (2017) Model-based design of process strategies for cell culture bioprocesses: state of the art and new perspectives. In: Gowder SJT (ed) New insights into cell culture technology. InTech
9. Puskeiler R, Kreuzmann J, Schuster C, Didzus K, Bartsch N, Hakemeyer C, Schmidt H, Jacobs M, Wolf S (2011) The way to a design space for an animal cell culture process according to Quality by Design (QbD). BMC Proc 5(Suppl 8):P12
10. Abu-Absi SF, Yang L, Thompson P, Jiang C, Kandula S, Schilling B, Shukla AA (2010) Defining process design space for monoclonal antibody cell culture. Biotechnol Bioeng 106:894–905
11. Möller J, Kuchemüller KB, Steinmetz T, Koopmann KS, Pörtner R (2019) Model-assisted design of experiments as a concept for knowledge-based bioprocess development. Bioprocess Biosyst Eng 42:867–882
12. Möller J, Hernández Rodríguez T, Müller J, Arndt L, Kuchemüller KB, Frahm B, Eibl R, Eibl D, Pörtner R (2020) Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. Comput Chem Eng 134:106693
13. Kuchemüller KB, Pörtner R, Möller J (2020) Efficient optimization of process strategies with model-assisted design of experiments. Methods Mol Biol 2095:235–249
14. Wu P, Ray NG, Shuler ML (1992) A single-cell model for CHO cells. Ann N Y Acad Sci 665:152–187
15. Möhler L, Flockerzi D, Sann H, Reichl U (2005) Mathematical model of influenza A virus production in large-scale microcarrier culture. Biotechnol Bioeng 90:46–58
16. López-Meza J, Araíz-Hernández D, Carrillo-Cocom LM, López-Pacheco F, Rocha-Pizaña MDR, Alvarez MM (2016) Using simple models to describe the kinetics of growth, glucose consumption, and monoclonal antibody formation in naive and infliximab producer CHO cells. Cytotechnology 68:1287–1300
17. Caramihai M, Severi I (2014) Bioprocess modeling and control. In: Matovic MD (ed) Biomass now – sustainable growth and use. InTech, Rijeka
18. Provost A, Bastin G (2004) Dynamic metabolic modelling under the balanced growth condition. J Process Control 14:717–728
19. Frahm B, Lane P, Atzert H, Munack A, Hoffmann M, Hass VC, Pörtner R (2002) Adaptive, model-based control by the open-loop-feedback-optimal (OLFO) controller for the effective fed-batch cultivation of hybridoma cells. Biotechnol Prog 18:1095–1103
20. Kern S, Platas-Barradas O, Pörtner R, Frahm B (2016) Model-based strategy for cell culture seed train layout verified at lab scale. Cytotechnology 68:1019–1032
21. Amribt Z, Niu H, Bogaerts P (2013) Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. Biochem Eng J 70:196–209
22. Möller J, Korte K, Pörtner R, Zeng A-P, Jandt U (2018) Model-based identification of cell-cycle-dependent metabolism and putative autocrine effects in antibody producing CHO cell culture. Biotechnol Bioeng 115:2996–3008
23. Kroll P, Hofer A, Stelzer IV, Herwig C (2017) Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering. Process Biochem 62:24–36
24. Möller J, Bhat K, Riecken K, Pörtner R, Zeng A-P, Jandt U (2019) Process-induced cell cycle oscillations in CHO cultures: Online monitoring and model-based investigation. Biotechnol Bioeng 116:2931–2943
25. Kalil SJ, Maugeri F, Rodrigues MI (2000) Response surface analysis and simulation as a tool for bioprocess design and optimization. Process Biochem 35:539–550
26. Costa AC, Atala DIP, Maugeri F, Maciel R (2001) Factorial design and simulation for the optimization and determination of control structures for an extractive alcoholic fermentation. Process Biochem 37:125–137

27. Parampalli A, Eskridge K, Smith L, Meagher MM, Mowry MC, Subramanian A (2007) Developement of serum-free media in CHO-DG44 cells using a central composite statistical design. Cytotechnology 54:57–68
28. Montgomery DC (2013) Design and analysis of experiments.8th edn. Wiley, Hoboken
29. Nasri Nasrabadi MR, Razavi SH (2010) Use of response surface methodology in a fed-batch process for optimization of tricarboxylic acid cycle intermediates to achieve high levels of canthaxanthin from Dietzia natronolimnaea HS-1. J Biosci Bioeng 109:361–368
30. Zhang H, Wang H, Liu M, Zhang T, Zhang J, Wang X, Xiang W (2013) Rational development of a serum-free medium and fed-batch process for a GS-CHO cell line expressing recombinant antibody. Cytotechnology 65:363–378
31. Horvath B, Mun M, Laird MW (2010) Characterization of a monoclonal antibody cell culture production process using a quality by design approach. Mol Biotechnol 45:203–206
32. Mandenius C-F, Graumann K, Schultz TW, Premstaller A, Olsson I-M, Petiot E, Clemens C, Welin M (2009) Quality-by-design for biotechnology-related pharmaceuticals. Biotechnol J 4:600–609
33. Mandenius C-F, Brundin A (2008) Bioprocess optimization using design-of-experiments methodology. Biotechnol Prog 24:1191–1203
34. Duvar S, Hecht V, Finger J, Gullans M, Ziehr H (2013) Developing an upstream process for a monoclonal antibody including medium optimization. BMC Proc 7
35. Legmann R, Schreyer HB, Combs RG, McCormick EL, Russo AP, Rodgers ST (2009) A predictive high-throughput scale-down model of monoclonal antibody production in CHO cells. Biotechnol Bioeng 104:1107–1120
36. Moran EB, McGowan ST, McGuire JM, Frankland JE, Oyebade IA, Waller W, Archer LC, Morris LO, Pandya J, Nathan SR, Smith L, Cadette ML, Michalowski JT (2000) A systematic approach to the validation of process control parameters for monoclonal antibody production in fed-batch culture of a murine myeloma. Biotechnol Bioeng 69:242–255
37. Dubey KK, Behera BK (2011) Statistical optimization of process variables for the production of an anticancer drug (colchicine derivatives) through fermentation: at scale-up level. New Biotechnol 28:79–85
38. Kleppmann W (2013) Versuchsplanung: Produkte und Prozesse optimieren.8th edn. Hanser, München
39. Myers RH, Anderson-Cook C, Montgomery DC (2016) Response surface methodology: process and product optimization using designed experiments. Wiley, Hoboken
40. Sandadi S, Ensari S, Kearns B (2006) Application of fractional factorial designs to screen active factors for antibody production by Chinese hamster ovary cells. Biotechnol Prog 22:595–600
41. Siebertz K, van Bebber D, Hochkirchen T (2010) Statistische Versuchsplanung: design of experiments (DoE). Springer, Berlin
42. Asghar A, Abdul Raman AA, Daud WMAW (2014) A comparison of central composite design and Taguchi method for optimizing Fenton process. TheScientificWorldJOURNAL 2014:869120
43. Del Castillo E (2007) Process optimization: a statistical approach. Springer, New York
44. Ferreira SLC, Bruns RE, Ferreira HS, Matos GD, David JM, Brandão GC, da Silva EGP, Portugal LA, dos Reis PS, Souza AS, dos Santos WNL (2007) Box-Behnken design: an alternative for the optimization of analytical methods. Anal Chim Acta 597:179–186
45. Goel T, Haftka RT, Shyy W, Watson LT (2008) Pitfalls of using a single criterion for selecting experimental designs. Int J Numer Methods Eng 75:127–155
46. Santner TJ, Williams BJ, Notz WI (2003) The design and analysis of computer experiments. Springer, New York
47. Lee GM, Kim EJ, Kim NS, Yoon SK, Ahn YH, Song JY (1999) Development of a serum-free medium for the production of erythropoietin by suspension culture of recombinant Chinese hamster ovary cells using a statistical design. J Biotechnol 69:85–93
48. Chun C, Heineken K, Szeto D, Ryll T, Chamow S, Chung JD (2003) Application of factorial design to accelerate identification of CHO growth factor requirements. Biotechnol Prog 19:52–57

49. Rouiller Y, Périlleux A, Vesin M-N, Stettler M, Jordan M, Broly H (2014) Modulation of mAb quality attributes using microliter scale fed-batch cultures. Biotechnol Prog 30:571–583
50. Yang WC, Lu J, Nguyen NB, Zhang A, Healy NV, Kshirsagar R, Ryll T, Huang Y-M (2014) Addition of valproic acid to CHO cell fed-batch cultures improves monoclonal antibody titers. Mol Biotechnol 56:421–428
51. Torkashvand F, Vaziri B, Maleknia S, Heydari A, Vossoughi M, Davami F, Mahboudi F (2015) Designed amino acid feed in improvement of production and quality targets of a therapeutic monoclonal antibody. PLoS One 10:e0140597
52. Ganguly J, Vogel G (2006) Process analytical technology (PAT) and scalable automation for bioprocess control and monitoring-A case study. Pharm Eng 26
53. Kreutz C, Timmer J (2009) Systems biology: experimental design. FEBS J 276:923–942
54. Smucker B, Krzywinski M, Altman N (2018) Optimal experimental design. Nat Methods 15:559–560
55. Walter É, Pronzato L (1997) Identification of parametric models from experimental data. Springer, London
56. Anselment B, Schoemig V, Kesten C, Weuster-Botz D (2012) Statistical vs. stochastic experimental design: an experimental comparison on the example of protein refolding. Biotechnol Prog 28:1499–1506
57. Banga JR, Balsa-Canto E (2008) Parameter estimation and optimal experimental design. Essays Biochem 45:195–209
58. Chaudhuri P, Mykland PA (1993) Nonlinear experiments: optimal design and inference based on likelihood. J Am Stat Assoc 88:538
59. Ford I, Titterington DM, Kitsos CP (1989) Recent advances in nonlinear experimental design. Technometrics 31:49
60. Franceschini G, Macchietto S (2008) Model-based design of experiments for parameter precision: state of the art. Chem Eng Sci 63:4846–4872
61. Moser A; Kuchemüller KB, Deppe S, Hernández Rodríguez T, Frahm B, Pörtner R, Hass VC, Möller J. Model-assisted DoE software: optimization of growth and biocatalysis in *Saccharomyces cerevisiae* bioprocesses. under reveision
62. Nargund S, Guenther K, Mauch K (2019) The move toward Biopharma 4.0. Genet Eng Biotechnol News 39:53–55
63. Möhler L, Bock A, Reichl U (2008) Segregated mathematical model for growth of anchorage-dependent MDCK cells in microcarrier culture. Biotechnol Prog 24:110–119
64. Shirsat NP, English NJ, Glennon B, Al-Rubeai M (2015) Modelling of mammalian cell cultures. In: Al-Rubeai M (ed) Animal cell culture, vol 9. Springer, Cham, pp 259–326
65. Pörtner R, Schäfer T (1996) Modelling hybridoma cell growth and metabolism — a comparison of selected models and data. J Biotechnol 49:119–135
66. Djuris J, Djuric Z (2017) Modeling in the quality by design environment: regulatory requirements and recommendations for design space and control strategy appointment. Int J Pharm 533:346–356
67. Berry B, Moretto J, Matthews T, Smelko J, Wiltberger K (2015) Cross-scale predictive modeling of CHO cell culture growth and metabolites using Raman spectroscopy and multivariate analysis. Biotechnol Prog 31:566–577
68. Pörtner R, Platas Barradas O, Frahm B, Hass VC (2016) Advanced process and control strategies for bioreactors. In: Current developments in biotechnology and bioengineering: bioprocesses, bioreactors and controls. Larroche C, Pandey A, Du G, Sanroman MA (eds) Elsevier Science: Saint Louis, 463–493
69. Shirsat N, Mohd A, Whelan J, English NJ, Glennon B, Al-Rubeai M (2015) Revisiting Verhulst and Monod models: analysis of batch and fed-batch cultures. Cytotechnology 67:515–530
70. Deppe S, Frahm B, Hass VC, Hernández Rodríguez T, Kuchemüller KB, Möller J, Pörtner R (2020) Estimation of process model parameters. Methods Mol Biol 2095:213–234
71. Storhas W (2013) Bioverfahrensentwicklung. Wiley, Weinheim

72. Hass VC, Pörtner R (2009) Praxis der Bioprozesstechnik: Mit virtuellem Praktikum. Spektrum Akad. Verl, Heidelberg
73. von Stosch M, Hamelink J-M, Oliveira R (2016) Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study. Bioprocess Biosyst Eng 39:773–784
74. Lei F, Rotbøll M, Jørgensen SB (2001) A biochemically structured model for Saccharomyces cerevisiae. J Biotechnol 88:205–221
75. Brüning S, Gerlach I, Pörtner R, Mandenius C-F, Hass VC (2017) Modeling suspension cultures of microbial and mammalian cells with an adaptable six-compartment model. Chem Eng Technol 40:956–966
76. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28:245–248
77. Lularevic M, Racher AJ, Jaques C, Kiparissides A (2019) Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions. Biotechnol Bioeng 116:2339–2352
78. Mantzaris NV, Daoutidis P, Srienc F (2001) Numerical solution of multi-variable cell population balance models. II. Spectral methods. Comput Chem Eng 25:1441–1462
79. Jandt U, Barradas OP, Pörtner R, Zeng A-P (2015) Synchronized mammalian cell culture: part II--population ensemble modeling and analysis for development of reproducible processes. Biotechnol Prog 31:175–185
80. Sanderson CS, Barford JP, Barton GW (1999) A structured, dynamic model for animal cell culture systems. Biochem Eng J 3:203–211
81. Jang JD, Sanderson CS, Chan LC, Barford JP, Reid S (2000) Structured modeling of recombinant protein production in batch and fed-batch culture of baculovirus-infected insect cells. Cytotechnology 34:71–82
82. Kontoravdi C, Wong D, Lam C, Lee YY, Yap MGS, Pistikopoulos EN, Mantalaris A (2007) Modeling amino acid metabolism in mammalian cells-toward the development of a model library. Biotechnol Prog 23:1261–1269
83. Jones B, Goos P (2012) I-optimal versus D-optimal split-plot response surface designs. J Qual Technol 44:85–101
84. Lawson J (2010) Design and analysis of experiments with SAS. CRC Press, Hoboken
85. Johnson RT, Montgomery DC, Jones BA (2011) An expository paper on optimal design. Qual Eng 23:287–301
86. Kenett R, Steinberg D (2007) New frontiers in the design of experiments. IEEE Eng Manag Rev 35:91
87. Steinberg DM, Lin DKJ (2006) Amendments and corrections. Biometrika 93:1025
88. Bursztyn D, Steinberg DM (2006) Comparison of designs for computer experiments. J Stat Plan Infer 136:1103–1119
89. Vining GG, Kowalski SM (2011) Statistical methods for engineers.3rd edn. Brooks/Cole Cengage Learning, Boston
90. Moser A. mDoE-toolbox
91. Beckmann TF, Krämer O, Klausing S, Heinrich C, Thüte T, Büntemeyer H, Hoffrogge R, Noll T (2012) Effects of high passage cultivation on CHO cells: a global analysis. Appl Microbiol Biotechnol 94:659–671
92. Wippermann A, Rupp O, Brinkrolf K, Hoffrogge R, Noll T (2015) The DNA methylation landscape of Chinese hamster ovary (CHO) DP-12 cells. J Biotechnol 199:38–46
93. Ulonska S, Kroll P, Fricke J, Clemens C, Voges R, Müller MM, Herwig C (2018) Workflow for target-oriented parametrization of an enhanced mechanistic cell culture model. Biotechnol J 13: e1700395
94. Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. J Geophys Res 106:7183–7192
95. Gagnon M, Hiller G, Luan Y-T, Kittredge A, DeFelice J, Drapeau D (2011) High-end pH-controlled delivery of glucose effectively suppresses lactate accumulation in CHO fed-batch cultures. Biotechnol Bioeng 108:1328–1337

96. Fu T, Zhang C, Jing Y, Jiang C, Li Z, Wang S, Ma K, Zhang D, Hou S, Dai J, Kou G, Wang H (2016) Regulation of cell growth and apoptosis through lactate dehydrogenase C over-expression in Chinese hamster ovary cells. Appl Microbiol Biotechnol 100:5007–5016
97. Andersen DC, Goochee CF (1995) The effect of ammonia on the O-linked glycosylation of granulocyte colony-stimulating factor produced by chinese hamster ovary cells. Biotechnol Bioeng 47:96–105
98. Ha TK, Kim Y-G, Lee GM (2015) Understanding of altered N-glycosylation-related gene expression in recombinant Chinese hamster ovary cells subjected to elevated ammonium concentration by digital mRNA counting. Biotechnol Bioeng 112:1583–1593

Adv Biochem Eng Biotechnol (2021) 177: 63–94
https://doi.org/10.1007/10_2020_151

Published online: 20 November 2020

Digital Twins for Bioprocess Control Strategy Development and Realisation

Christian Appl, André Moser, Frank Baganz, and Volker C. Hass

Contents

C. Appl and V. C. Hass (✉)
Faculty of Medical and Life Sciences, Furtwangen University, Villingen-Schwenningen, Germany

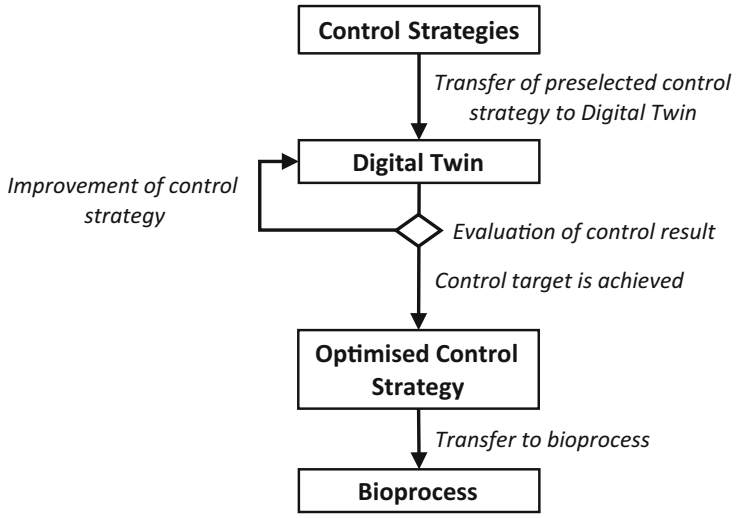
Department of Biochemical Engineering, University College London, London, UK
e-mail: Christian.Bernhard.Appl@hs-furtwangen.de; christian.appl.18@ucl.ac.uk; Volker.Hass@hs-furtwangen.de; v.hass@ucl.ac.uk

A. Moser
Faculty of Medical and Life Sciences, Furtwangen University, Villingen-Schwenningen, Germany
e-mail: Andre.Moser@hs-furtwangen.de

F. Baganz
Department of Biochemical Engineering, University College London, London, UK
e-mail: f.baganz@ucl.ac.uk

Abstract New innovative Digital Twins can represent complex bioprocesses, including the biological, physico-chemical, and chemical reaction kinetics, as well as the mechanical and physical characteristics of the reactors and the involved peripherals. Digital Twins are an ideal tool for the rapid and cost-effective development, realisation and optimisation of control and automation strategies. They may be utilised for the development and implementation of conventional controllers (e.g. temperature, dissolved oxygen, etc.), as well as for advanced control strategies (e.g. control of substrate or metabolite concentrations, multivariable controls), and the development of complete bioprocess control. This chapter describes the requirements Digital Twins must fulfil to be used for bioprocess control strategy development, and implementation and gives an overview of research projects where Digital Twins or "early-stage" Digital Twins were used in this context. Furthermore, applications of Digital Twins for the academic education of future control and bioprocess engineers as well as for the training of future bioreactor operators will be described. Finally, a case study is presented, in which an "early-stage" Digital Twin was applied for the development of control strategies of the fed-batch cultivation of *Saccharomyces cerevisiae*.

Graphical Abstract Development, realisation and optimisation of control strategies utilising Digital Twins



Keywords
Bioprocess, Control strategy development, Digital Twin, Operator training simulator (OTS)

Abbreviations

AMBC	Advanced and model-based control
CHO	Chinese hamster ovary (mammalian cell)
DCU	Digital control unit
DLL	Dynamic link library
DO	Dissolved oxygen
DoE	Design of experiment
EtOH	Ethanol
GUI	Graphical user Interface
MPC	Model predictive control
NMPC	Nonlinear model predictive control
OLFO	Open-loop-feedback-optimal strategy
OTS	Operator training simulator
P	Product (Ethanol)
P	Proportional (P-controller)
P&ID	Piping and instrumentation diagram
PCS	Process control system
PI	Proportional integral (PI-controller)
PID	Proportional integral derivate (PID-controller)
RQ	Respiratory quotient
S	Substrate (Glucose)
SSF-BC	Simultaneous saccharification, fermentation, and biocatalysis
STR	Stirred tank reactor
X	Dry biomass density (*S. cerevisiae*)

1 Introduction

The development of control strategies for bioprocesses poses huge challenges for process engineers. The need for new tools that can help with this task, therefore, is enormous. Optimisation of controllers during production runs is usually exceedingly difficult or even impossible. Thus, bioprocess operation must be interrupted for control optimisation. Interruptions of a production run, as well as inadequate control, can lead to immense financial losses, which must be avoided. A promising approach to this issue is the application of Digital Twins. The development or optimisation of control strategies may be performed using this tool, thus leading to a shortened start-up time for the newly developed or optimised bioprocess control scheme.

In the early 2000s, the Digital Twin concept was first applied in mechanical engineering [1–3]. Digital Twins are often seen as virtual representations of physical systems and can map the entire life cycle of the physical system [2]. Various authors already published definitions of the term Digital Twin [1–5]. This chapter and *Chapter: Moser, Appl, Brüning, Hass "Mechanistic Mathematical Models as a*

Basis of Digital Twins for process optimization", which is also in this book series are mainly based on the definition given by El Saddik [3]:

> *Digital twins are (. . .) digital replications of living as well as non-living entities that enable data to be seamlessly transmitted between the physical and virtual worlds.*

For further explanations refer to *Chapter: Moser, Appl, Brüning, Hass "Mechanistic Mathematical Models as a Basis of Digital Twins for process optimization"*, which is also in this book series.

This chapter covers Digital Twins for the development, optimisation and realisation or implementation of bioprocess control strategies on a real process that correspond to the Digital Twin definition given by El Saddik [3], as well as operator training simulators (OTSs), which are considered by the authors to be "early-stage" Digital Twins. Although OTSs are mainly used for training purposes, they also offer enormous potential for bioprocess development, similarly to Digital Twins. OTSs are usually adapted to the real process during development or when there are significant changes in the real process.

In the last section of this chapter, a case study is presented where an "early-stage" Digital Twin was used to develop process control strategies for the fed-batch cultivation of *Saccharomyces cerevisiae* (*S. cerevisiae*) in a stirred tank reactor (STR).

2 Advanced Bioprocess Control Development, Realisation and Optimisation Using Digital Twins

Initial approaches for the application of Digital Twins as a tool for control strategy development have been successfully established in the chemical industry [4–7]. Due to the recognised potential, the application of Digital Twins as a tool for the development of control strategies is also gaining increasing interest for bioprocesses.

Within this chapter, the suitability of Digital Twins for the development, optimisation and realisation of bioprocess control strategies will be highlighted. First, the general approach when using Digital Twins for the development of control strategies is outlined. Subsequently, the requirements that Digital Twins must fulfil to be used as a tool for the development of control strategies and which challenges control engineering must overcome in the case of bioprocess control are described. Finally, in the presented case study, application examples for the utilisation of Digital Twins for bioprocess control strategy development are described.

2.1 General Approach

In the author's opinion, the quality of Digital Twins is of utmost importance for the development of control and automation strategies [8]. The basis of applicable Digital

Twins is a dynamic mathematical model, which can map the biological, chemical and physical phenomena of the real process in detail [9]. This dynamic mathematical process model should be coupled to a graphical user interface (GUI) [9]. Users can monitor and make changes to the virtual process using graphical icons in the GUI. From the author's point of view, it is advantageous, if the structure of the GUI corresponds to the process control system (PCS) on the physical counterpart. The Digital Twin GUI is a functional image, derived from the P&ID (piping and instrumentation diagram) flow chart of the real bioprocess and thus, also serves as a realistic replica of important parts of the control and automation model. A realistic GUI of a Digital Twin can, therefore, be used to check the usability (including typical operating errors), as well as the control and automation of the real bioprocess. The model of a Digital Twin is parameterised based on real process data to represent the behaviour of the physical process [10]. Another possibility to keep Digital Twin and the real process as identical as possible is an online and at-line data connection between the "twins". This enables the adaption of the Digital Twin using online and at-line data, which is particularly useful if the real process frequently changes its characteristics.

During process development or optimisation, Digital Twins can be used for the following applications:

1. Determination of suitable controller types.
2. Improvement of controller performance.
3. Improvement of the overall process performance through appropriate process control strategies.

If, for example, suitable controllers (e.g. for temperature, dissolved oxygen or product concentration) should be designed, the controller type can be selected based on simulations with the Digital Twin. An early step in controller selection should be the definition of appropriate control targets [8]. When controlling the temperature of a bioreactor, such control targets are, e.g., a short rise time, a high control accuracy (especially important for temperature-sensitive organisms, particularly mammalian cells) or a low overshoot. For example, the conventional proportional integral derivative (PID) control can be compared to a more complex nonlinear model predictive control (NMPC) by applying them to a Digital Twin. If both control strategies yield equally good control results, PID control would be preferred, because it is cheaper and easier to handle.

Once a control strategy has been able to control the virtual process satisfactorily, the results are transferred to the real process. The transfer of the developed control strategy from the Digital Twin to the real process may be further simplified if the Digital Twin and the real process are linked to the identical PCS [8].

To illustrate the general approach of process control design utilising a Digital Twin, the case study in Sect. 4 presents the selection and optimisation of suitable control strategies for the cultivation of *S. cerevisiae*.

2.2 Design of Digital Twins as Control Strategy Development Tools

To utilise a Digital Twin for the development of both conventional (e.g. single loop PID control) and advanced control (e.g. multivariable controllers, model predictive control), it must fulfil specific requirements that have to be considered during the design process of the Digital Twin. According to Hass [11], desirable characteristics of a functionally useful Digital Twin include realistic simulation of the biological, physical and chemical processes, accurate representation of automation and control actions and a GUI with a similar "look and feel" to that of the real plant [11]. Mathematical models used in Digital Twin development are classified broadly as mechanistic, non-mechanistic or hybrid models [9, 10, 12]. In this context, a model refers to a mathematical representation of certain aspects of a real-world object or phenomenon. Non-mechanistic models use sets of experimental data to represent observed phenomena by fitting parameters based on the available datasets. Mechanistic models seek to represent experimental observations based on the underlying biological, chemical, and physical mechanisms occurring in the system. Mechanistic models offer excellent predictive capabilities beyond the original experimental conditions used for model development. By contrast, non-mechanistic models only offer very restricted predictive capabilities [2, 9–12]. Mathematical modelling for a Digital Twin involves several key steps. The first step is a definition of the process using appropriate diagrams and charts. A process flow diagram and a piping and instrumentation diagram (P&ID) are excellent starting points for system definition [10, 13, 14]. Ideally, verbal process description and expected modelling targets including levels of model accuracy are specified at this stage. Following system definition, appropriate mathematical models that sufficiently describe the physical, biological, and chemical processes in the system are formulated based on literature research [9, 14]. To structure the process model, it has been suggested to divide the model into smaller sub-models. One approach is the shell model introduced by Blesgen et al. [15, 16] and extended by Hass et al. [17]. In this case, the overall mathematical model of the Digital Twin is divided into a biological sub-model, physico-chemical sub-model, a reactor sub-model, a plant and peripheral sub-model as well as a control and automation sub-model (see also *Chapter: Moser, Appl, Brüning, Hass "Mechanistic Mathematical Models as a Basis of Digital Twins for process optimization"*, which is also in this book series). Depending on the requirements of the Digital Twin, the shell model can be extended or reduced in complexity.

2.2.1 Software Tools for the Design of Digital Twins

Further steps in Digital Twin development include model implementation using suitable tools, model parameterisation and finally model validation using experimental data. Several modelling tools for the development of Digital Twins are readily available and easy to use, but they do not provide the flexibility and

adaptability needed to model all aspects of bioprocesses, as they were originally designed for modelling of chemical processes. With the increasing focus on bioprocess development, significant effort has been invested in the development of model libraries for bioprocess unit operations in recent years. Software systems for parameter estimation and computation of algebraic and differential equations provide a user-friendly and adaptable environment for model development and implementation of Digital Twins [9–11].

For the design of Digital Twins or "early-stage" Digital Twins that can be used for the development, optimisation and realisation of control strategies, there are already a variety of software packages available. Table 1 lists a selection of vendors and associated software products and summarises the most important features of the respective software packages. Most of the Digital Twin development tools listed are designed for the chemical industry (e.g. UniSim Competency Suite [18] or IndissPlus [19]), but some are also suitable for the development of bioprocess Digital Twins (e.g. WinErs/C-eStIM [20, 21], PerceptiveAPC [22] or TMODS [23]).

2.3 Control Strategies for Bioprocesses

The multi-phase system in a bioprocess sets highest demands on measurement and control technology [32–34]. To maintain optimal conditions for the entire process, the composition of the liquid phase (e.g. medium), the suspended gas phase (e.g. oxygen, carbon dioxide) and the dispersed solid phase (e.g. cells, cell assemblies, enzymes) must be monitored continuously [32]. Furthermore, complex dynamics showing a wide range of time constants make it difficult to control the process without sufficient process knowledge [32]. For example, the induction of a gene through a temperature shift or the addition of a chemical inducer affects the process several minutes after the expression of the desired protein because the formation of a metabolically active protein will cause a time delay. This kind of knowledge must be available and utilised for successful bioprocess control based on detailed process analytics [32–35].

The choice of control strategies mainly depends on the selected bioprocess and the available reactor type [33, 34]. In general, controllers are divided according to continuous (e.g. PID control, soft sensor control) and discontinuous behaviour (e.g. model predictive control (MPC) or nonlinear model predictive control (NMPC)) [34]. Controllers with continuous behaviour calculate and transmit continuous control signals based on the current process characteristics [34]. Among the best-known continuous controllers are the "conventional" controllers like two-point-, three-point-, proportional- (P-), proportional-integral- (PI-) or PID-controllers. Controllers with discontinuous behaviour only calculate control signals or profiles at specific process points [34].

As an example, conventional control strategies such as PI or PID control are generally used to control temperature [34]. In many cases, the control system should be able to maintain the desired setpoint, due to the rather weak influence of

Table 1 Digital twin development tools for the process industry (adapted from [10])

Vendor	Software package	Key features (according to the vendors)
Aspen Technology	Aspen OTS Framework [24]	▪ Data communication links handle the exchange of data and commands ▪ User interfaces support different views of the application for operators, engineers and training instructors
DuPont Industrial Biosciences	TMODS [23]	▪ Fully customised to match plant configuration, conditions, compositions, control schemes, safety interlocks and GUIs
Honeywell	UniSim Competency Suite [18]	▪ Customisable framework for a structured operator competency management system ▪ Interactive, navigable, panoramic 2D field operator training environment based on high-resolution photographs of the facility
Ingenieurbüro Dr.-Ing. Schoop GmbH	WinErs/C-eStIM [20, 21]	▪ Modular process automation system ▪ Provides a flexible, process control and simulation system suitable for industrial, didactical and research applications ▪ Complete process monitoring and operation via a user-editable GUI ▪ Simple graphical editing of controls and simulations via block structures, logic plans and GRAFCET with no prior programming knowledge required
Wood Group (John Wood Group)	ProDyn [25]	▪ Offers off-the-shelf and customer-specific solutions ▪ Operator training and learning systems, abnormal situation management, and process troubleshooting ▪ Can be used to develop and test plant procedures
NovaTech	NovaTech Ethanol Training Simulator, D/3 DCS [26]	▪ Allows breweries, biofuels facilities, and other process plants to develop real-to-life plant simulations ▪ Training on complex process control techniques and correcting behavioural patterns ▪ Trend visualisation, process analytics and control loop performance monitoring and optimisation
Outotec	HSC Sim [27]	▪ Various simulation and modelling applications based on independent chemical reactions and process units ▪ Graphical flowsheet and spreadsheet type process unit models
Perceptive Engineering	PerceptiveAPC [22]	▪ Tools for monitoring, analysis or predictive control, in a logical, intuitive interface, for both batch and continuous processes ▪ Training module and easy-to-use templates to tune and validate the right controller (also model-predictive control (MPC)) for the process

(continued)

Table 1 (continued)

Vendor	Software package	Key features (according to the vendors)
Protomation BV	Protomation OTS [28]	▪ A real-time dynamic model that covers the complete operating window ▪ Allows accurate simulation and training in the entire operating range of the plant (from start-up conditions up to normal operation and upset conditions)
CORYS	IndissPlus [19]	▪ Models based on first principles of chemical engineering with rigorous thermodynamics calculation and physical component properties database ▪ Can accurately represent plant start-up and shutdown, in addition to a variety of design and abnormal operating conditions
Siemens	SIMIT OTS [29]	▪ Based on the dynamic modelling of the plant ▪ Flexible modelling is possible, the process can be emulated as a whole or in parts
SimGenics	SimuPACT [30]	▪ The integrated software platform enables engineers to develop high fidelity, full-scope power and process plant simulators ▪ Intuitive GUI which allows engineering analysis and operator training on the same simulation platform
Yokogawa	Yokogawa OTS [31]	▪ OTS constantly synchronises with the plant control system ▪ Able to predict plant internal states and plant responses, contributing to optimised plant operations

disturbances. More complex processes, such as the enzymatic hydrolysis of lignocellulosic biomass, can be significantly improved by using advanced temperature control. In this process, endoglucanase and exoglucanase are used, which show a different temperature optimum. If model-based temperature control is applied in this case, enzyme-specific temperature gradients can be operated, reducing the consumption of enzymes and significantly increasing the yield of the desired product [36].

Table 2 lists common control variables (e.g. temperature, pH value or dissolved oxygen (DO)) of bioprocesses with their most used control strategies.

Simple control tasks can be treated using conventional controllers. For more demanding control tasks, such as concentration control, the use of advanced and model-based control strategies such as MPC or NMPC has been suggested [34, 35, 48, 49]. The choice of suitable control strategies is not only dependent on the controlled variable. If, for example, DO control is considered, on-off feedback, PID control or more complex model-based control like MPC is used depending on the requirements. In the subsequent sections, some advanced control strategies will be described that may be developed and tuned utilising Digital Twins.

Table 2 Control strategies for key variables in bioprocesses

Control variable	Applied control strategy
Temperature	PI control [34], MPC [36], NMPC [37]
pH	PI control [38]
DO	On-Off-Feedback control [34], PID control [34], Cascade Control [38], MPC [34]
Flow rate (nutrient media, etc.)	PI control [38]
Pressure	PI control [38]
Concentration (substrate, product, etc.)	PI control [39], fuzzy control [40], NMPC [41–43], OLFO [44–47]

2.3.1 Advanced and Model-Based Control Strategies

Advanced and model-based control strategies (AMBC) like NMPC are of great interest in the case of processes with fast dynamics because these controllers reduce the response time [34]. They do not operate just based on the current state of the system, instead, the control action is based on the calculated evolution of the system. AMBCs utilise integrated mathematical process models for the prediction of future process behaviour. At the end of each sampling period, the future course of the control trajectory is optimised using a process model [34]. The control trajectory that fulfils the chosen optimisation criterion best is then applied to the real process [34].

The use of AMBC has already been investigated for different bioprocesses in several research works. For fermentations of *S. cerevisiae* NMPC was used to maximise the ethanol (EtOH) yield by controlling the glucose solution feed rate [42]. For the fed-batch cultivation of Chinese hamster ovary (CHO) mammalian cells, a glucose concentration fixed setpoint control was implemented and tuned to enhance product quality and reduce costs [43]. To enhance the sugar concentration in a cellulose hydrolysation process in a stirred tank reactor, NMPC was applied to control the feed rates of substrate and cellulase enzymes solutions [50]. Furthermore, temperature and humidity gradients of solid-state fermentation were controlled by NMPC [51].

In all listed research works the use of AMBC resulted in higher product concentrations at lower resource demands as compared to processes with conventional control strategies.

2.3.2 Open-Loop-Feedback-Optimal (OLFO) Control Strategy

A special form of AMBC is the open-loop-feedback-optimal (OLFO) strategy [52, 53]. The OLFO controller belongs to the class of adaptive NMPCs. It consists of a process model, a model parameter identification part, and an optimisation part (see Fig. 1). Model parameters are estimated frequently based on available online and/or offline data. The updated model parameters are passed on to the optimisation

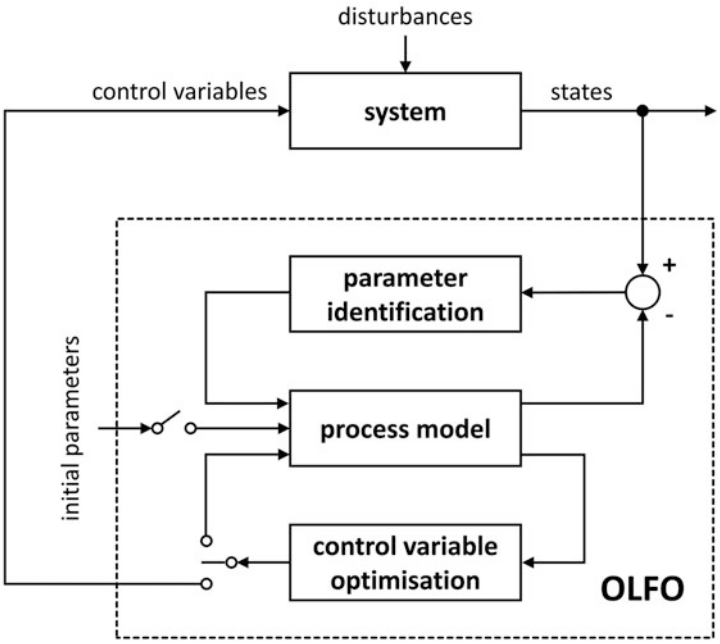


Fig. 1 Structure of the Open-Loop-Feedback-Optimal (OLFO) control strategy [53]

part, where process trajectories like substrate feeding profiles are calculated. Several optimisation criteria, such as maximised product concentrations, may be implemented in the controller. The OLFO control strategy has been investigated in a receding horizon [8, 53] and a moving horizon version [45, 47] for bioprocesses.

The OLFO strategy is particularly superior to other process control strategies if the processes are in an early development phase and have not yet been optimised. The performance of the OLFO algorithm for suspension cell cultures has already been demonstrated by Witte et al. [53], Frahm et al. [45–47] and Li et al. [44]. In the case study presented in Sect. 4.2.3 the application of the OLFO control strategy for fed-batch cultivation of *S. cerevisiae* will be explained in more detail.

2.4 Digital Twin Based Development, Realisation and Optimisation of Control Strategies for Bioprocesses

In the early to mid-1980s, first OTSs representing "early-stage" Digital Twins were used for operator training in the chemical, nuclear and energy industries. In the late 1980s and early 1990s, the implementation of OTSs in the chemical industry evolved from pioneering work to common practice [54]. Today, Digital Twins are widely used in industries with high capital investment, complex processes and

severe consequences of plant or operator failure such as the offshore oil and gas industry [7, 54, 55]. Older educational facilities for training in the oil and gas industry were based on physical copies of the control room, which are expensive and no longer needed [54]. Almost simultaneously with the first appearance of Digital Twins in the chemical industry, they were used as a tool for control strategy development [54]. In the beginning, these were relatively simple control engineering tasks, but they became more complex with the advancing development of Digital Twins [54, 55].

Dudley et al. (2008) [7] described the use of a Digital Twin of a pebble bed modular reactor plant for the development and testing of control strategies before using them on the real plant. He et al. (2019) [4] described the use of a Digital Twin for the Tennessee Eastman benchmark process. Effectiveness and performance of the Digital Twin in the development of control strategies were demonstrated in the presence of realistic fault scenarios. Three types of process faults, i.e., sensor faults, actuator faults and process disturbances were investigated and the corresponding fault size and temporal behaviour were discussed. All simulation studies and numerical results indicated that the proposed configurations are valid for safe operations in the event of a process fault. Zhang et al. (2019) [6] described the use of a Digital Twin for carbon emission reduction in intelligent manufacturing. Here, the plants' carbon emission is predicted by the Digital Twin model. A carbon emission control strategy was then optimised utilising the Digital Twin, to minimise exhaust gas emissions.

Compared to chemical processes, the application of Digital Twins for bioprocesses is still in its infancy. Thoroughness is required for modelling bioprocesses since a wide variety of parallel reactions take place at the same time. Even small changes of key process variables, such as pH or temperature, may have an immense influence on the kinetics [33].

Pörtner et al. (2011) used an "early-stage" Digital Twin for the optimisation of process control strategies for mammalian cell cultivations [56]. The developed bioprocess simulator is a digital replica of the cultivation of mammalian cell lines in a small scale STR. The bioprocess simulator was used to simulate the impact of various constant feed rates of glucose and glutamine during fed-batch on cell density and antibody concentration of a mammalian cell line. The feed rates were determined by design of experiments (DoE) methods. By using the bioprocess simulator, the cultivation process could be optimised in a considerably shorter time and fewer experiments compared to process control optimisation on the real process.

In a contribution by Hass et al. [17] the utilisation of an industrial biotechnology OTS was presented. Control strategies that were developed using a new bioethanol plant OTS illustrated the potential for enhanced resource efficiency and reduced energy consumption. According to the authors, the potential savings in raw materials have a direct impact on the long-term profitability of the bioethanol plant and enable a reduction of operating costs. By using the OTS, the time course and dynamics of the entire plant could be analysed and subsequently optimised using new process control strategies. Performing such a study on a real plant would have been overly complex and expensive, if not impossible.

3 Digital Twins as Training and Educational Tools

Digital Twins or "Digital Twin-like" simulators may also be used in industry to train reactor and plant operators and in academia to educate future control and process engineers. In this context, Digital Twins are usually referred to as OTSs [9–11, 57].

OTSs became increasingly popular since the mid-twentieth century, for the use in various sectors, including the chemical and related industries [10, 54]. The reason was the increasing complexity of process engineering plants with sophisticated automation and process control strategies placing enormous demands on the skills of the process operators [10, 54]. Several papers were published reviewing the development and use of OTSs in the chemical process industry [54, 58, 59].

OTSs offer the possibility to train future reactor operators and bioprocess engineers in a very practical way without carrying out the real process. Even actions to compensate process malfunctions may be trained safely. Impairments on ongoing production processes due to training are avoided. OTSs can be described as "early-stage" Digital Twins.

The development and use of OTSs particularly for bioprocesses are beginning to attract increasing academic interest [10]. Several research groups have investigated the applications of OTSs for bioprocesses. The common premise of the presented research works confirms experiences from the chemical industry. Model-based OTSs are an efficient means to improve the training experience of students and to increase plant operators skills in handling complex bioprocesses [13, 14, 16, 60, 61].

Table 3 gives an overview of already existing OTSs for bioprocesses.

Hass et al. [17] developed one of the earliest OTSs for a complex biorefinery process. OTSs were created for the bioethanol fermentation and the distillation process. Also, a separate biomass power plant training simulator was developed. The mathematical process models were created and implemented using the FORTRAN programming language [65]. The process control software WinErs [20] was used to link process control and the simulation models. PCS-like GUIs were developed to obtain full operator training simulators. Functions were implemented to simulate the processes at different speeds depending on the desired training target. The different OTSs were designed for the training of students as well as industrial operators in the handling of biorefineries and biomass power plants. Encouraging training outcomes were reported [10, 17].

A research project by Gerlach et al. [61] presented an OTS for the training of bioengineering students and plant operators on the operational procedures and production skills required in recombinant protein production processes. To enable the model to accurately represent the complex relations of factors in a recombinant protein production process, the authors outlined that several metabolic interactions affecting biomass yield, productivity and cellular viability need to be mapped in the OTS model. To maintain numerical efficiency, a trade-off between model complexity and accuracy had to be found by capturing the most important metabolic processes in the OTS model, without the model being cumbersome and numerically

Table 3 OTS applications for bioprocesses and biorefineries [10]

Application	Development tools	Validation	Reference
Conceptual design of 2-step biodiesel synthesis process (theoretical 120,000 t per year capacity biorefinery)	Aspen plus dynamics Aspen OTS framework	Unknown	Ahmad et al. [62]
30 L jacketed batch reactor hydrodynamic and thermal behaviour parameterisation	Unisim design	Simulated temperature profiles compared with laboratory reactor temperature measurements	Balaton et al. [63]
Anaerobic biogas production in a 10 L laboratory reactor	FORTRAN (biological and physico-chemical sub-models) WinErs (reactor and plant sub-models, plus automation, process control and GUI)	Experimental data from literature validated with simulation results	Blesgen and Hass [16]
Bioethanol production from *S. cerevisiae* (15 L STR) and green fluorescence protein production using *E. coli* (6 L fed-batch bioreactor)	Biological and physico-chemical models integrated into WinErs as dynamic link libraries (DLLs)	Substrate consumption, product formation and biomass yields were compared between laboratory reactor and simulator runs	Gerlach et al. [57]
Large-scale commercial bioethanol process (reactors ranging in size from 30,000 L to 280,000 L)	Process models written in C++ were implemented as DLLs in WinErs	Model validation not presented	Gerlach et al. [14]
Integrated cultivation and homogenisation for recombinant protein production in a 10 L STR	Process models written in C++ were implemented as DLLs in WinErs	Substrate consumption, product formation and biomass yields were compared between laboratory reactor and simulator runs	Gerlach et al. [64]
Integrated wastewater biodegradation and membrane filtration in a 10 L submerged membrane bioreactor (SMBR)	The biological model was written and implemented in Pascal, while process automation and GUI were developed using Delphi 2009	Experimental data from literature validated with simulation runs	González Hernández et al. [60]
Describes the development of a coding framework combined with a commercial process control software for rapid process model development in chemical and biochemical engineering	eStIM coding framework used for biological and process model development and WinErs is used for automation and process control	Experimental data from *S. cerevisiae* production compared with simulation results	Hass et al. [65]
Bioethanol production, crossflow filtration and rectification column	Process models written in C++ were implemented as DLLs in WinErs.	Laboratory fermenter, membrane filtration unit and distillation runs were	Hass et al. [17]

(continued)

Table 3 (continued)

Application	Development tools	Validation	Reference
(15 L laboratory bioreactors used for EtOH production)	GRAFCET used for developing automation sequences	used to validate simulator runs	
Mammalian cell line cultivation with the production of antibodies in 2 L laboratory bioreactors	Process models written in FORTRAN were implemented as DLLs in WinErs	Experimental data from mammalian cell line cultivation compared with simulation results	Pörtner et al. [56]

difficult to calculate. The effectiveness of OTS training for the education of bioengineering students was evaluated with promising results [10, 61].

Another possible application of OTSs is their use for training in the context of control engineering. Currently, training in control engineering is frequently theoretical and abstract, since investigations of different control strategy behaviour in real processes are difficult, time and cost-intensive and the number of available plants for training is limited. With the help of Digital Twins or other simulation tools, a wide variety of control strategies may be investigated in a short time and their impact on bioprocess performance can be demonstrated. In future, applications of OTSs will become even more diverse. New control strategies may be tested first on the OTSs. This guarantees safe operation of the real plant. Furthermore, full plant process control and operation strategies may be developed and optimised based on OTSs or Digital Twins.

4 Case Study

The objective of this case study, which is based on a work of Appl et al. [8], is to demonstrate the methodology and advantages of Digital Twins for the development of bioprocess control strategies using a fed-batch cultivation of *S. cerevisiae* as an illustrative example. Two process control strategies (respiratory quotient (RQ) feedback control and OLFO control) were developed and optimised using the "early-stage" Digital Twin "Simultaneous saccharification and fermentation simulator" (SSF-BC-Simulator). The target for both control strategies was to maximise the dry biomass concentration (*S. cerevisiae*) in a cultivation time of 48 h.

4.1 Digital Twin "SSF-BC-Simulator"

The Digital Twin "SSF-BC-Simulator" is a further development of the "BioProzessTrainer" [33, 66]. It is used to train bioengineering students for the operation of bioprocesses as well as a control strategy development tool.

The Digital Twin can map the starch hydrolysis, the cultivation of *S. cerevisiae* and the whole-cell biocatalysis of ethyl (S)-3-hydroxybutyrate from ethyl acetate in a small scale STR (Biostat C, 20 L, B. Braun). The development of the "SSF-BC Simulator" was carried out using the procedure described in Sect. 2.2. The integrated dynamic mathematical model was written in C++ and was implemented in WinErs [20, 65]. Using the Digital Twin, it is possible to accelerate the simulation of the bioprocesses up to 100-fold. The Digital Twin can be monitored and operated via the GUI shown in Fig. 2.

The GUI in Fig. 2 presents the process equipment (e.g. reactor, feed tanks, etc.) as well as all measured value displays (e.g. temperature, pH value, DO, etc.) and all essential functions of the control system (e.g. temperature or DO control) to the user of the Digital Twin. Behind each measured value display or control button, sub-models represent the real measuring or control instrument. The reactor properties and the biological process are mapped in the dynamic mathematical model of the Digital Twin. The GUI is part of the control and automation model within the Digital Twin. To use the Digital Twin for the development, optimisation and realisation of control strategies, it is therefore important that the GUI corresponds to the PCS of the real process with high similarity.

4.1.1 Parameterisation of the Digital Twin "SSF-BC-Simulator"

For the parameterisation of the dynamic mathematical process model implemented in the Digital Twin "SSF-BC-Simulator", a variety of parameterisation experiments were carried out, using batch and fed-batch cultivations.

The procedure of model parameterisation will be illustrated using a dataset from a laboratory experiment where an aerobic fed-batch cultivation was carried out in a small scale STR (Biostat C, 20 L, B. Braun). The temperature was controlled at 30°C, the pH value at 4.5 and the DO at 10%. At the beginning of the cultivation, a nutrient medium was supplied in the STR (Batch medium). After the batch phase of the cultivation, a fed-batch nutrient medium was fed to the STR (see Table 4).

During the cultivation process, the following state variables required for process monitoring and process control were measured (see Table 5).

After the experiment was carried out, the model of the Digital Twin "SSF-BC-Simulator" was parameterised using the Nelder-Mead simplex algorithm, written in R [67], to adjust the values of selected parameters to match the simulated with the measured data satisfactorily.

Figures 3 and 4 present the measured state variables of the fed-batch *S. cerevisiae* cultivation in a small scale STR compared to the simulated time courses of the Digital Twin (after parameterisation).

Figure 3 shows that in the batch phase of the experiment (0–7 h), glucose was consumed. Ethanol (EtOH) was formed, which was subsequently metabolised again (diauxic growth). The biomass density shows a slight increase during the batch phase. After the substrate feed has been activated (7–25 h), the dry biomass concentration increases to a value of more than 30 g L^{-1}. At a processing time of

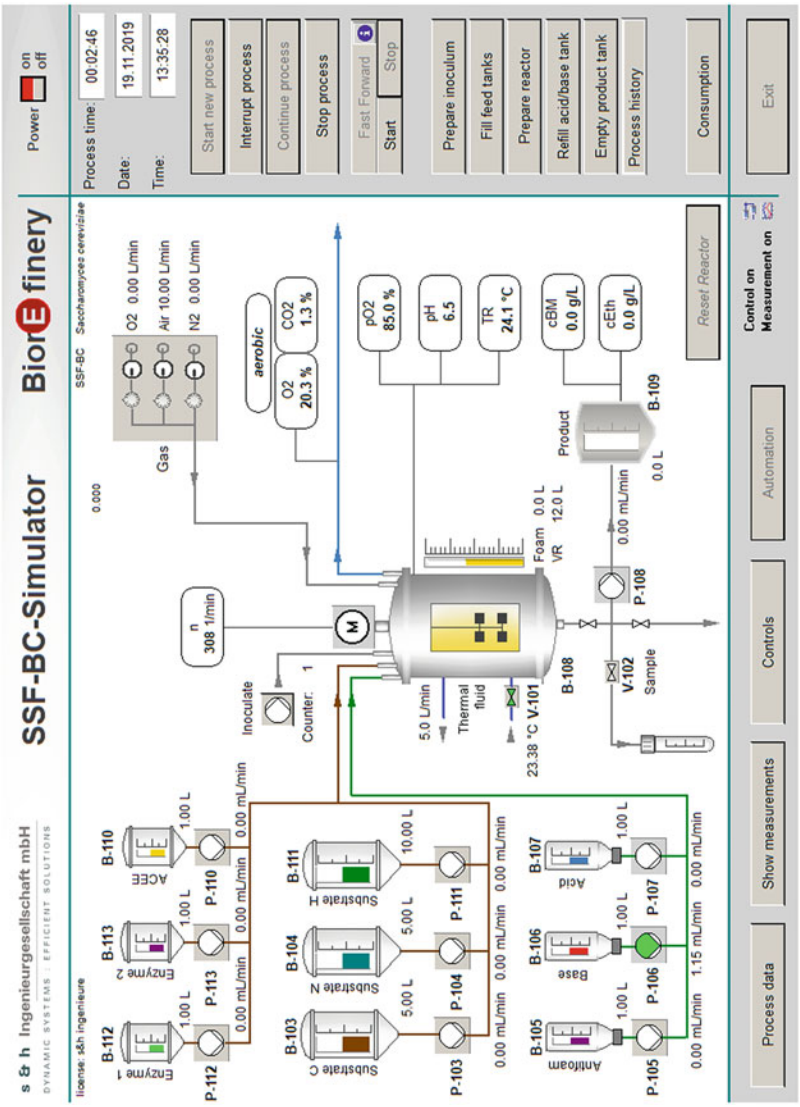


Fig. 2 GUI of the early-stage Digital Twin "SSF-BC-Simulator" [20], with illustrations, e.g. STR, tanks, pumps or sampling vessels that represent the real process, display windows, e.g. temperature, pH value or DO to monitor the virtual process and buttons to set, e.g., simulation speed, start conditions or stirrer speed

Table 4 Nutrient media composition

Component	Batch medium (g L^{-1})	Fed-batch medium (g L^{-1})
Glucose	5.0	300
Yeast extract	0.6	40
Peptone from soy	0.6	40
Ammonium sulphate	0.6	40

Table 5 Measured state variables during the parameterisation experiment

Measured state variable	Abbreviation	Unit
Substrate (glucose) concentration	S	g L^{-1}
Product (EtOH) concentration	P	g L^{-1}
Dry biomass (*S. cerevisiae*) concentration	X	g L^{-1}
Fed-batch medium feed rate	$Feed_S$	mL min^{-1}
Oxygen in the exhaust gas	O_2	%
Carbon dioxide in the exhaust gas	CO_2	%

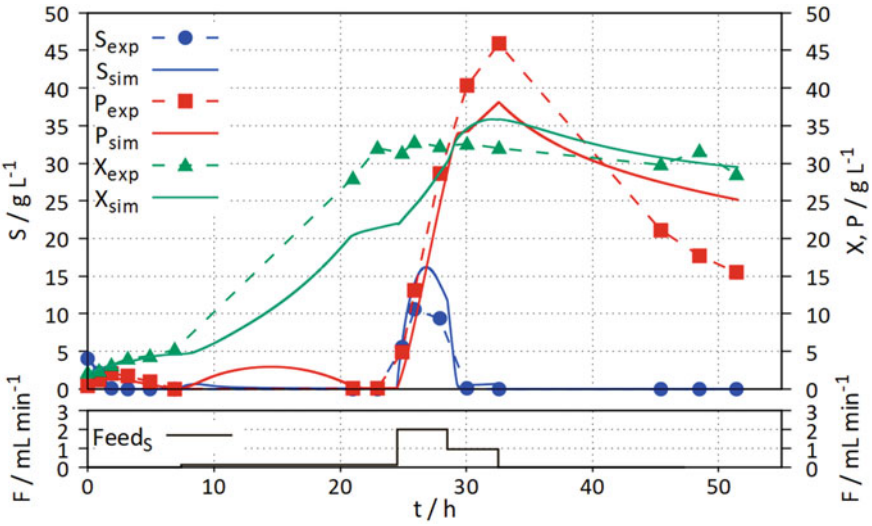


Fig. 3 Comparison of measured data (exp) from a small scale STR with simulation results (sim), S: substrate (glucose) P: product (EtOH), X: dry biomass concentration (*S. cerevisiae*). The bottom figure shows substrate feed profile

25 h, the substrate feed was increased by a factor of almost 10, which resulted in an increase of the glucose concentration to more than 10 g L^{-1}. An increase in the ethanol concentration to more than 45 g L^{-1} was observed, due to the Crabtree effect. The high ethanol concentration inhibited the growth of *S. cerevisiae* and the dry biomass concentration stagnated at a level of 30 g L^{-1}. After the substrate feed has been reduced, the glucose concentration decreased to nearly 0 g L^{-1}, followed

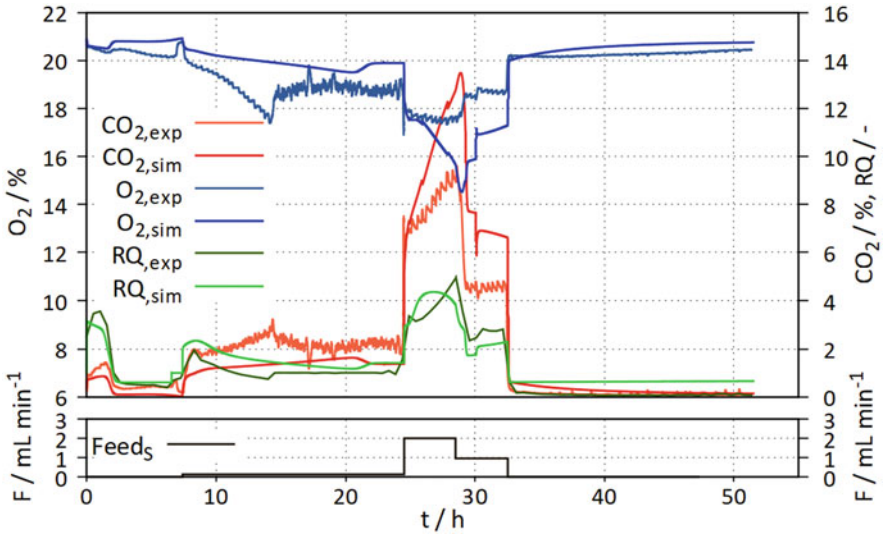


Fig. 4 Comparison of measured exhaust gas data (CO_2, O_2) and calculated RQ values from a small scale STR experiment (exp) with simulation results (sim). The bottom figure shows substrate feed profile

by ethanol consumption down to a concentration of 15 g L^{-1}. However, after 22 h of process time, no further biomass growth could be observed.

In Fig. 4 it can be seen that these effects are also reflected in the measured exhaust gas values. Special attention should be paid to the course of the RQ value (see Sect. 4.2.2 for details). At the beginning of the batch phase (0–3 h), the RQ rises to a value above 3, indicating ethanol formation. After the initial phase, the RQ value dropped below 1, now indicating ethanol consumption. At the beginning of substrate feeding, a parallel increase in CO_2 formation and O_2 consumption can be observed, thus indicating good aerobic growth of *S. cerevisiae*. During this phase, the RQ settled at a value around 1.0. From a processing time of 25 h, the substrate feed was strongly increased. In this period a large increase in CO_2 formation can be seen, however, the consumption of O_2 increases only slightly, leading to an RQ value of above 3. This high RQ value again indicates the formation of ethanol, which is confirmed by the offline ethanol concentration measurements. At the end of the cultivation, both the formation of CO_2 and the O_2 consumption value dropped close to zero, indicating weak metabolism and poor growth. These observations confirm that particularly the RQ-value is a valuable indicator for various metabolic effects as also stated previously [68].

4.1.2 Digital Twin "SSF-BC-Simulator" for the Development of Control Strategies

To ensure that the Digital Twin is suitable for the development of control strategies for the cultivation of *S. cerevisiae*, it must be able to represent the time courses of the experimental data described in Figs. 3 and 4. These time courses do not have to be simulated exactly, but the associated effects must be reproduced. For the development of the RQ feedback control strategy utilising the Digital Twin, it is important that exhaust gas measurements, RQ value time course and associated effects can be mapped. For the development of the OLFO controller with the Digital Twin, it is necessary to simulate the course of the concentrations of substrate, product and biomass and the corresponding effects.

Figure 3 shows that the time course of the measured variables can be mapped by the Digital Twin with a high agreement. Also, ethanol formation due to the Crabtree effect can be represented by the Digital Twin (0–3 h and 25–33 h). It is also clearly recognisable that high ethanol concentrations inhibit the growth of the cultivated *S. cerevisiae* strain in the simulation (30–52 h).

Figure 4 illustrates that the time courses of the measured exhaust gas values can almost be exactly reproduced by the Digital Twin. Also, in the simulation, an increase in the RQ value occurs if ethanol is formed due to the Crabtree effect (0–3 h and 25–33 h). Furthermore, at the end of the simulated cultivation, almost no CO_2 is formed or O_2 is consumed, corresponding to a low growth rate.

The results presented in Figs. 3 and 4 illustrate the high potential of the Digital Twin for the development of an RQ feedback control strategy and an OLFO strategy for the cultivation of *S. cerevisiae*. In the presented study, the control target was to maximise the dry biomass concentration (*S. cerevisiae*). To achieve this target, it is important to dose the substrate feed in such a way that the cells are sufficiently supplied with glucose. However, overdosing substrate may lead to ethanol formation (Crabtree effect), which then might cause growth inhibition.

4.2 *Digital Twin Based Development of Control Strategies for the Cultivation of* S. cerevisiae

During process control strategy development, the different strategies were first applied to the "SSF-BC-Simulator". Simulations with varying controller designs and tunings were then carried out on the Digital Twin until the desired controller performance was achieved. Afterwards, the experimental validation of the control strategies on the real plant took place. If the control result was still unsatisfactory, further controller improvements were tested using the Digital Twin, before validating the controllers on a real cultivation process. By using the Digital Twin, many complex experiments in the STR with elaborate preparation, execution and analysis could be avoided in the development of the control strategies, which resulted in a

resource-saving of over 50%. Also, the acceleration mode of the Digital Twin offered a significant reduction in development time.

4.2.1 Experimental Setup

To realise a smooth transfer of the control strategies between the "twins", the Digital Twin and the small scale STR were connected to the identical process control system WinErs [20], in which also the controllers were implemented (see Fig. 5).

Since both, the real STR and the Digital Twin were connected to the identical PCS, the control strategies could be quickly and variably applied and transferred to the real and simulated process. Both the PCS and the control strategies (RQ feedback and OLFO) were realised in separate coupled WinErs projects, which leads to high compatibility.

4.2.2 Development of Respiratory Quotient (RQ) Feedback Control for the Cultivation of *S. cerevisiae*

The RQ feedback control strategy is an established soft sensor control strategy used for fed-batch cultivations of *S. cerevisiae* [68]. To ensure optimal growth of *S. cerevisiae* the RQ should be kept close to a value of 1.0. For the determination of the RQ value, the composition of the exhaust gas from the reactor during the cultivation is measured using a gas analyser (SIDOR, Sick). The RQ value can be calculated from the measured mole fractions of O_2 and CO_2 in the supply air and the exhaust gas (Eqs. 1–3),

$$y_{i,0} = 1 - \left(y_{O_2,0} + y_{CO_2,0}\right) \tag{1}$$

$$y_{i,1} = 1 - \left(y_{O_2,1} + y_{CO_2,1}\right) \tag{2}$$

$$RQ = \frac{\left(y_{CO_2,1} \cdot \left(\frac{y_{i,0}}{y_{i,1}}\right)\right) - y_{CO_2,0}}{y_{O_2,0} - \left(y_{O_2,1} \cdot \frac{y_{i,0}}{y_{i,1}}\right)} \tag{3}$$

where $y_{i,0}$ is the mole fraction of inert components in the supply air, $y_{i,1}$ is the mole fraction of inert components in the exhaust gas, $y_{O_2,0}$ is the mole fraction of O_2 in the supply air (assumption: 0.2096), $y_{O_2,1}$ is the mole fraction of O_2 in the exhaust gas, $y_{CO_2,0}$ is the mole fraction of CO_2 in the supply air (assumption: 0.00035) and $y_{CO_2,1}$ is the mole fraction of CO_2 in the exhaust gas.

To realise the RQ feedback control strategy a PI controller was chosen. Based on the difference between the RQ value and RQ setpoint, the PI controller calculated the appropriate substrate feed and transmitted it to the bioreactors digital control unit (DCU) every 5 min.

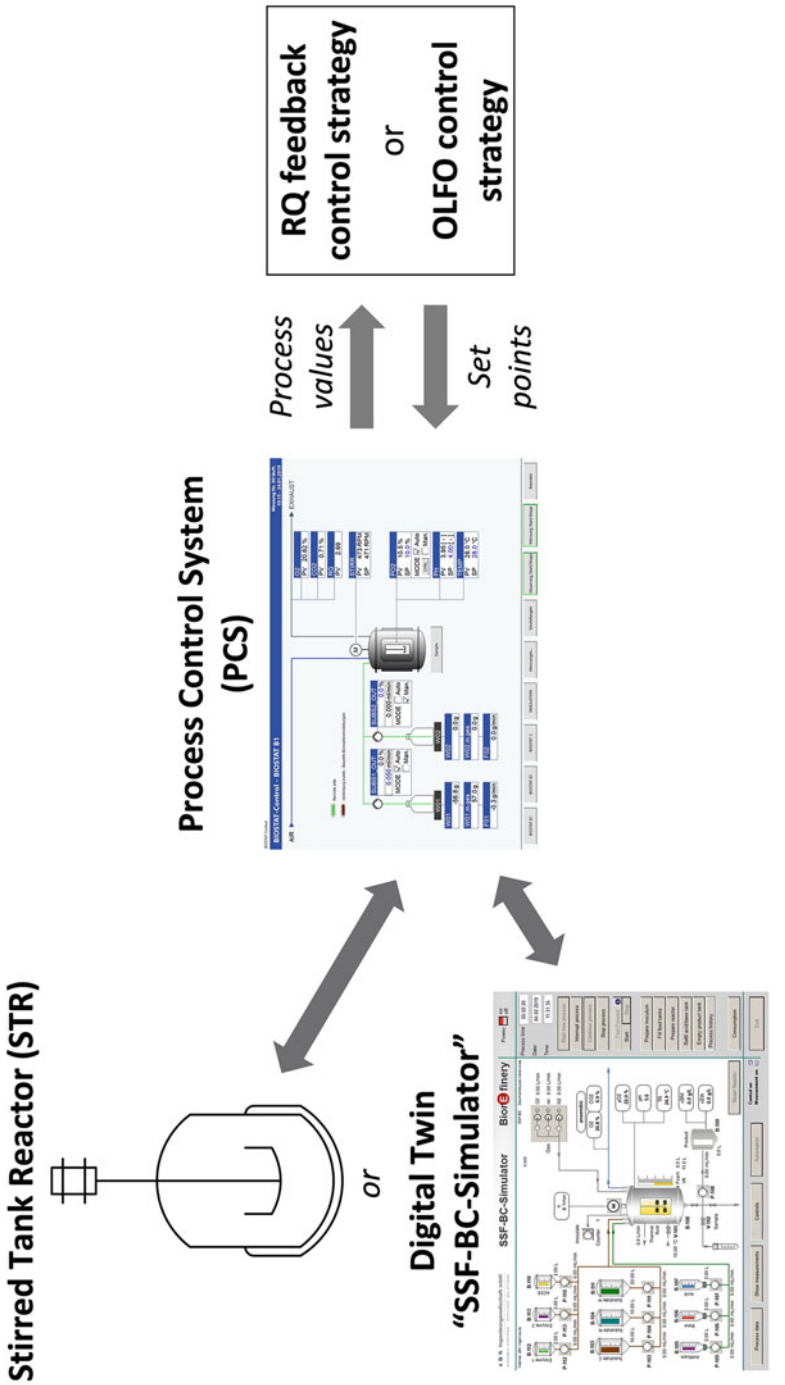


Fig. 5 Linking of STR, Digital Twin and PCS (with associated control strategies) in the Digital Twin based development of control strategies for the cultivation of *S. cerevisiae*

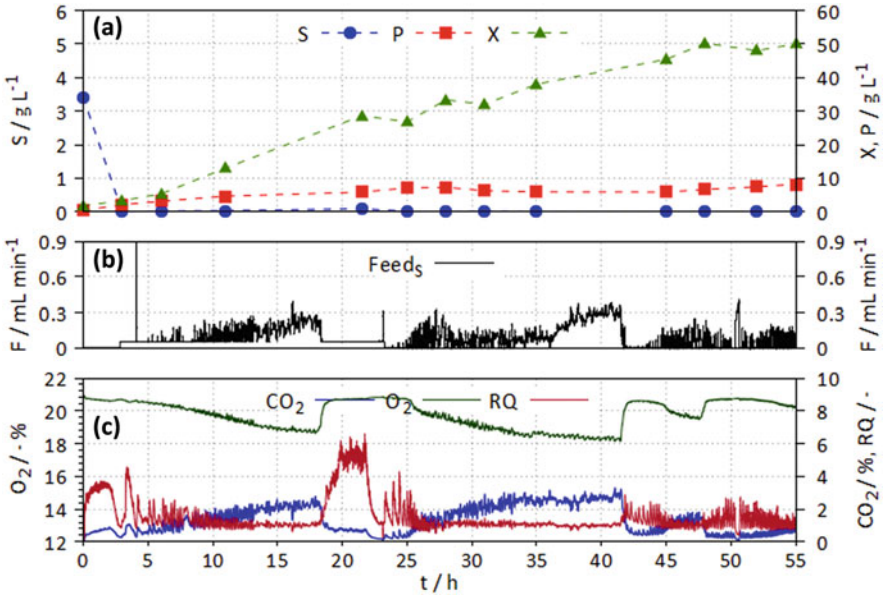


Fig. 6 Results of an RQ feedback-controlled *S. cerevisiae* cultivation in a STR

In the development process of the RQ feedback control strategy on the Digital Twin, various RQ value setpoints were tested, the controller parameters (gain, integration time) of the PI controller were adjusted and the transfer intervals of the calculated substrate feed rates to the DCU were varied. Furthermore, different ratios of glucose and nitrogen sources in the feed medium were investigated. To achieve the predetermined control target of 50 g L^{-1} after a processing time of 48 h, four simulations on the Digital Twin were performed.

The transfer of the RQ feedback control strategy to the real process took place after simulations on the Digital Twin yielded a dry biomass concentration of more than 50 g L^{-1} within 48 h. Then, the RQ feedback control strategy was experimentally validated on the real cultivation process in the small scale STR. The results of the real RQ feedback-controlled cultivation of *S. cerevisiae* in a small scale STR are presented in Fig. 6.

Figure 6b shows that the substrate feed ($Feed_S$) started at 3 h. At this time, the batch phase was finished and the RQ Feedback controller was switched on. After that, the mean substrate feed rate increased steadily up to 18 h. The addition of nutrient medium leads to a steady increase in dry biomass concentration up to 25 g L^{-1} (Fig. 6a). Figure 6c shows that both, O_2 consumption and CO_2 formation, increase during the first 18 h. The resulting RQ value stabilises to a value close to 1.1. After a processing time of 18 h, the RQ value increased to a value of up to 6, resulting in a substrate feed rate, controlled to the set minimum value of 0.05 ml min^{-1}. When the substrate was depleted, the RQ value dropped below 1.1 again (approx. 25 h), the substrate feed rate started to increase. At processing times

of 43 h and 47 h, the same effect observed at 18 h can be seen in an attenuated form. One explanation for the sudden increase in the RQ value is the composition of the nutrient medium. Among other components, yeast extract was used as a nitrogen source, which contains high amounts of both nitrogen and carbon. The fraction of residual yeast extract in the medium was rather high, leading to an accumulation of carbon sources and thus to an increasing RQ value due to the Crabtree effect. In the Digital Twin model, the carbon component in the nitrogen sources was not considered, which is why this effect could only be recognised in the real experiment. Despite this limitation of the Digital Twin model, an RQ feedback control could be developed based on the Digital twin, leading to more than 50 g L^{-1} dry biomass concentration in the real process, with less than 10 g L^{-1} ethanol produced within 48 h.

It took about 2 days to develop the RQ feedback control for the cultivation of *S. cerevisiae* on the Digital Twin (simulations, controller adaptations). Real cultivation of 48 h in an STR, including preparation and evaluation, is expected to take about 1 week. If, instead of the simulations on the Digital Twin, real cultivations had to be carried out during the control strategy development process, the development time would have been extended to up to 3 weeks. Besides the significant time savings, the consumption of resources (nutrient media components, energy, etc.) was also significantly reduced due to the reduced number of real cultivations.

4.2.3 Development of Open-Loop-Feedback-Optimal (OLFO) Control for the Cultivation of *S. cerevisiae*

The principle of the OLFO control strategy has been described in Sect. 2.3.2. The suitability of the "SSF-BC-Simulator" as a tool for the development of the OLFO control strategy for the cultivation of *S. cerevisiae* was illustrated in Fig. 3, Sect. 4.1.

The core of the OLFO controller is a relatively simple mathematical model for the cultivation of *S. cerevisiae*, which is different from the process model within the presented Digital Twin. The controller model is limited to map the consumption of glucose and nitrogen, the growth of *S. cerevisiae* and the formation of the side product ethanol. The mathematical OLFO controller model was adapted based on either measured (real process) or simulated (Digital Twin) concentrations of substrate (glucose), product (ethanol) and biomass density (*S. cerevisiae*). In the optimisation part of the OLFO controller, substrate feed rate trajectories were optimised at several points during the real or simulated (Digital Twin) process using the adapted mathematical process model, where the adaption was based on the data available up to the actual processing time. The substrate feed rate trajectory yielding the highest concentration of dry biomass at the end of the simulated cultivation (OLFO process model) was transferred to the PCS at each time point of model adaption and process optimisation.

During controller development using the Digital Twin, six simulations were carried out in total. After each simulation, the simulated cultivation results were evaluated and the control strategy was adjusted to approach the control target

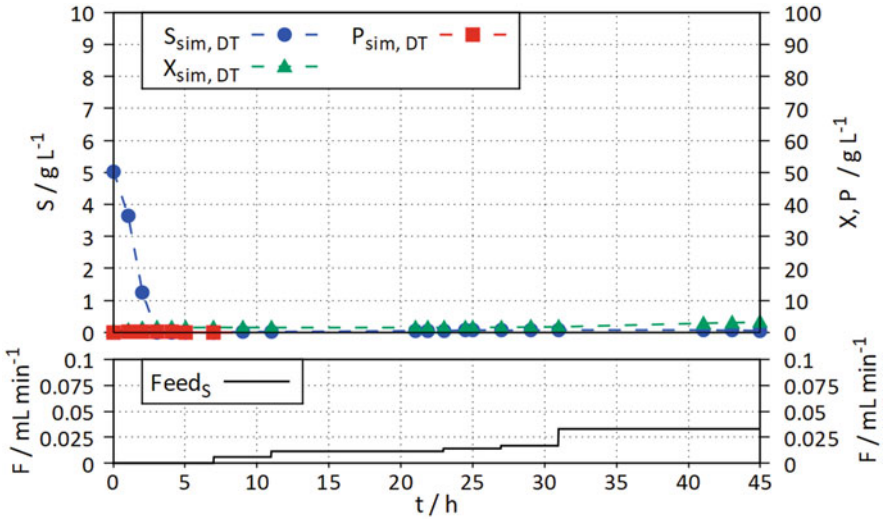


Fig. 7 Result of the first OLFO controlled simulated cultivation of *S. cerevisiae* using the Digital Twin "SSF-BC-Simulator"

(50 g L^{-1} dry biomass concentration within 48 h). The result of the first OLFO controlled simulated cultivation of *S. cerevisiae* is presented in Fig. 7.

In the first OLFO controlled Digital Twin cultivation of *S. cerevisiae,* only a low dry biomass concentration of 4 g L^{-1} could be achieved within the processing time of 48 h, due to low substrate feed rates (max. 0.03 mL min^{-1}) determined by the OLFO controller. A detailed analysis revealed an ethanol inhibition in the mathematical process model already starting at less than 5 g L^{-1}. Consistently, the OLFO controller calculated low substrate feed rates to avoid ethanol formation. However, the resulting low glucose concentration limited growth.

Increasing the ethanol inhibition constant in the mathematical process model to approx. 30 g L^{-1} led to an increase in the final simulated dry biomass concentration (15 g L^{-1}). However, the set control target could not yet be achieved. Based on subsequent simulations with the Digital Twin, further controller model adjustments such as modifying the metabolic rates related to the Crabtree effect, adjustments of uptake rates, etc. were performed. The intervals for model adaptation and subsequent substrate feed optimisations were varied and different compositions of the nutrient medium were examined via simulations with the Digital Twin.

In the sixth OLFO controlled cultivation simulated on the Digital Twin, the set control target eventually was exceeded by reaching a final biomass density of 80 g L^{-1} within 48 h (Fig. 8) and less than 10 g L^{-1} ethanol.

The resulting OLFO controller (developed on the Digital Twin) was transferred to the real process for experimental validation. Figure 9 shows the results of the OLFO controlled *S. cerevisiae* cultivation.

In the OLFO controlled real cultivation, a dry biomass concentration of more than 50 g L^{-1} was achieved within 48 h. Both, the substrate feed rates (Fig. 9b) and the

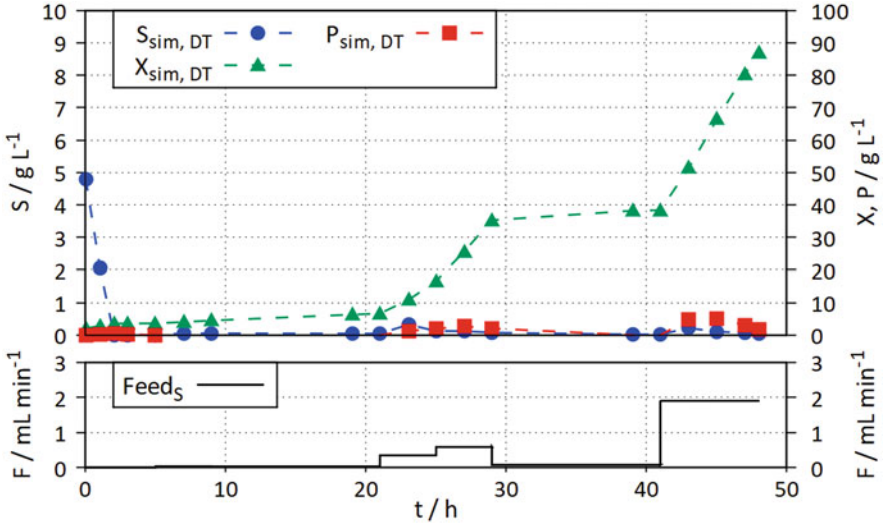


Fig. 8 Sixth OLFO controlled cultivation of *S. cerevisiae* on the Digital Twin "SSF-BC-Simulator"

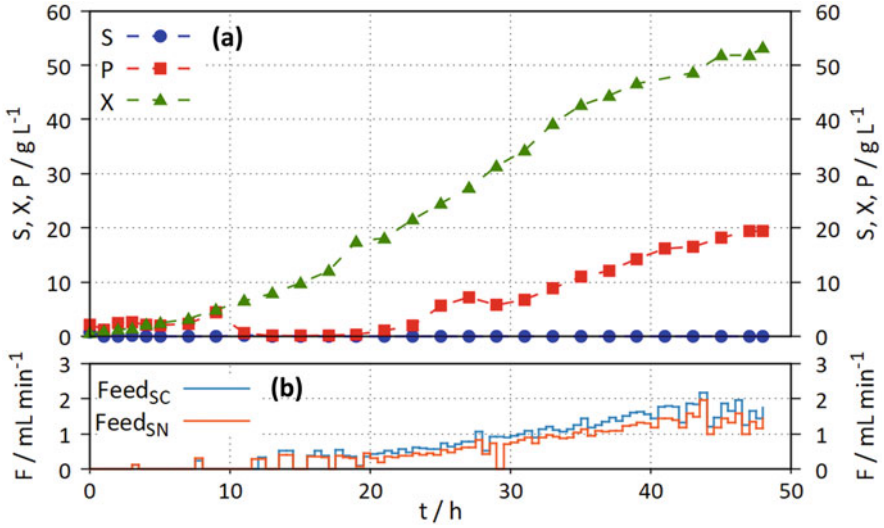


Fig. 9 OLFO controlled *S. cerevisiae* cultivation in a 20 L STR (Biostat C, B. Braun)

dry biomass concentration (Fig. 9a) increase steadily over the entire process time. The ethanol concentration never exceeded 20 g L^{-1}.

It took about 2 weeks to develop the OLFO control for the cultivation of *S. cerevisiae* on the Digital Twin (simulations, controller adaptations). Real cultivation of 48 h in an STR, including preparation and evaluation, is expected to take

about 1 week. If, instead of the simulations on the Digital Twin, real cultivations had to be carried out during the control strategy development process, the development time would have been extended to up to 2 months. Besides the significant time savings, the consumption of resources was also significantly reduced due to the smaller number of real cultivations.

4.2.4 Case Study Discussion

This case study demonstrated the enormous potential of the Digital Twin "SSF-BC-Simulator" to support the control strategy development and optimisation for the cultivation of *S. cerevisiae*. By utilising the Digital Twin, it was possible to effectively develop both control that uses online values (RQ feedback control) and control that uses offline values (OLFO control). By conducting simulations using the Digital Twin, real experiments could be avoided that would have been associated with the consumption of resources and time. By using the Digital Twin, an estimated amount of resources of about 60% and time of about 50% could be saved in the development process of both control strategies compared to conventional control strategy development.

In this case study, we were able to demonstrate the beneficial utilisation of Digital Twins for the development, optimisation and realisation of bioprocess control strategies. An important prerequisite for the Digital Twin utilisation for control development is the validation of a high accuracy in mapping the bioprocess dynamics.

The presented Digital Twin "SSF-BC-Simulator" is also capable of mapping the enzymatic process of starch hydrolysis as well as the biocatalysis of ethyl (S)-3-hydroxybutyrate. For these processes various control strategies will be developed in future, supported by the Digital Twin.

5 Conclusion and Future Perspectives

This chapter demonstrated the enormous potential of Digital Twins or "early-stage" Digital Twins as a control strategy development tool and their application to bioprocesses. The use of Digital Twins enables the development of advanced controllers that increase the efficiency of bioprocesses. By accelerated and parallel running simulations on the Digital Twin, the development time is drastically reduced compared to conventional control strategy development. In the past, production usually had to be interrupted to investigate the dynamic behaviour of the bioprocessing plant under consideration, as well as the dynamics of different controlled systems, which is necessary for the development of control strategies. By using Digital Twins, the production plants can remain in operation during controller development and optimisation. The presented case study demonstrates a rapid and effective controller transfer to the real plant as soon as the new controllers have been

successfully developed utilising the Digital Twin. An ideal process operation not only requires well-designed and tuned controllers but also well-trained plant operators. This can be achieved using OTSs that may be considered as "early-stage Digital Twins". From the further development of OTSs, educational Digital Twins have emerged, which are characterised by the following features:

1. High fidelity representation of biological, physico-chemical and chemical kinetics.
2. Detailed technical simulation of the reactor environment including peripheral equipment.
3. Realistic investigation of various control strategies.
4. Accelerated and resource-saving simulation (digital experimentation and training).

As new advanced bioprocessing plants are put into operation worldwide, the challenge of covering the need for suitably qualified operators to run these plants will increase. Educational Digital Twins are an effective tool to meet this challenge. In the future, simple and cost-effective educational Digital Twin development tools are required to adequately handle the additional complexities present in bioprocesses.

Acknowledgements The authors would like to thank C. Fittkau and S. Dreßler for their excellent laboratory work at Furtwangen University. We gratefully appreciate that parts of the presented work have been funded by the German Federal Ministry of Education and Research, Innovation Alliance prot P.S.I. (FKZ: 031B0405C).

References

1. Grieves M (2016) Origins of the digital twin concept: working paper
2. Glaessgen E, Stargel D (2012) The digital twin paradigm for future NASA and U.S. Air Force Vehicles: 22267B. https://doi.org/10.2514/6.2012-1818
3. El Saddik A (2018) Digital twins: the convergence of multimedia technologies. IEEE Multi-Media 25:87–92. https://doi.org/10.1109/MMUL.2018.023121167
4. He R, Chen G, Dong C et al (2019) Data-driven digital twin technology for optimized control in process systems. ISA Trans 95:221–234. https://doi.org/10.1016/j.isatra.2019.05.011
5. Zobel-Roos S, Schmidt A, Mestmäcker F et al (2019) Accelerating biologics manufacturing by modeling or: is approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin? PRO 7:94. https://doi.org/10.3390/pr7020094
6. Zhang C, Ji W (2019) Digital twin-driven carbon emission prediction and low-carbon control of intelligent manufacturing job-shop. Proc CIRP 83:624–629. https://doi.org/10.1016/j.procir.2019.04.095
7. Dudley T, de Villiers P, Bouwer W et al (2008) The operator training simulator system for the pebble bed modular reactor (PBMR) plant. Nucl Eng Des 238:2908–2915. https://doi.org/10.1016/j.nucengdes.2007.12.028
8. Appl C, Fittkau C, Moser A et al (2019) Adaptive, model-based control of Saccharomyces cerevisiae fed-batch cultivations. In: AIDIC SERVIZI SRL (ed) Book of abstracts: bridging science with technology, pp 1504–1505

9. Hass VC, Kuhnen F, Schoop K-M (2005) Rapid design of interactive operator-training simulators for training and education. In: 7th world congress of chemical engineering, WCCE 2005, 10th-14th July
10. Isimite J, Baganz F, Hass VC (2018) Operator training simulators for biorefineries: current position and future directions. J Chem Technol Biotechnol 93:1529–1541. https://doi.org/10.1002/jctb.5583
11. Hass VC (2016) Operator training simulators for bioreactors. In: Mandenius C-F (ed) Bioreactors: design, operation and novel applications, vol 69. Wiley, Weinheim, pp 453–486
12. Pavé A (2012) Modeling living systems: from cell to ecosystem. In: Environmental engineering series. ISTE Wiley, London
13. Hass VC, Knutzsch S, Gerlach I et al (2012) Towards the development of a training simulator for biorefineries. Chem Eng Trans:247–252. https://doi.org/10.3303/CET1229042
14. Gerlach I, Hass V, Mandenius C-F (2015) Conceptual design of an operator training simulator for a bio-ethanol plant. PRO 3:664–683. https://doi.org/10.3390/pr3030664
15. Blesgen A (2009) Entwicklung und Einsatz eines interaktiven Biogas-Echtzeit-Simulators. Dissertation, Universität Bremen
16. Blesgen A, Hass VC (2010) Efficient biogas production through process simulation †. Energy Fuel 24:4721–4727. https://doi.org/10.1021/ef9012483
17. Hass VC, Kuntzsch S, Schoop K-M et al. (2014) Resource efficiency studies using a new operator training simulator for a bioethanol plant. In: PRES 2014, 17th conference on process integration, modelling and optimisation for energy saving and pollution reduction: PRES 2014, 23–27 August 2014, Prague, Czech Republic. AIDIC Associazione Italiana di Ingegneria Chimica ČSCHI Česká Společnost Chemického Inženýrství, Milano, pp 541–546
18. Honeywell (2020) UniSim competency suite. https://www.honeywellprocess.com/en-US/explore/products/advanced-applications/unisim/unisim-competency-suite/Pages/default.aspx. Accessed 18 Aug 2020
19. CORYS (2020) Indiss Plus®. https://www.corys.com/en/indiss-plusr. Accessed 18 Aug 2020
20. Ingenieurbüro Dr.-Ing.Schoop GmbH (2018) WinErs: process control and automation system on PC under Windows, Hamburg, Germany
21. Hass VC, Kuhnen F, Schoop K-M (2005) An environment for the development of operator training systems (OTS) from chemical engineering models. Comput Aided Chem Eng:289–293. https://doi.org/10.1016/S1570-7946(05)80170-1
22. Perceptive Engineering (2020) PerceptiveAPC - key features and tools. https://www.perceptiveapc.com/software/features/. Accessed 18 Aug 2020
23. DuPont Industrial Biosciences (2020) Operator training simulator and training solutions for STRATCO® alkylation - DuPont industrial biosciences. http://cleantechnologies.dupont.com/technologies/stratcor/stratcor-equipment-services/alkylation-technology-training-solutions/. Accessed 18 Aug 2020
24. Aspen Technology (2008) Aspen OTS framework: best-in-class technology to configure and build operator training simulator applications. https://www.aspentech.com/uploadedfiles/products/templates/aspen_ots.pdf. Accessed 18 Aug 2020
25. Wood (2018) ProDyn - operator training simulator software. https://www.woodplc.com/capabilities/digital-and-technology/software,-applications-and-analytics/prodyn-operator-training-simulator-software. Accessed 18 Aug 2020
26. NovaTech (2017) Training simulators. NovaTech Process Control and Optimization. https://www.novatechweb.com/process-control/training-simulators/. Accessed 18 Aug 2020
27. Outotec (2020) HSC Sim: process simulation module. https://www.outotec.com/products-and-services/technologies/digital-solutions/hsc-chemistry/hsc-sim-process-simulation-module/. Accessed 18 Aug 2020
28. Protomation (2019) Custom made OTS. https://protomation.com/custom-made-ots/. Accessed 18 Aug 2020

29. Siemens AG (2020) SIMIT Simulation. https://new.siemens.com/global/de/produkte/automatisierung/industrie-software/simit.html. Accessed 18 Aug 2020
30. SimGenics (2020) SimuPACT. https://www.simgenics.com/page/simupact. Accessed 18 Aug 2020
31. Yokogawa (2020) Operator training simulator (OTS) which supports to acquire plant operation skills by using it with a dynamic virtual plant model. https://www.yokogawa.com/solutions/solutions/energy-management/operator-training-simulator/. Accessed 18 Aug 2020
32. Hitzmann B, Scheper T (2018) Bioprozessanalytik und -steuerung. In: Chmiel H, Takors R, Weuster-Botz D (eds) Bioprozesstechnik. Springer, Berlin, pp 263–294
33. Hass VC, Pörtner R (2011) Praxis der Bioprozesstechnik: Mit virtuellem Praktikum, 2. Aufl. Spektrum Akad. Verl., Heidelberg
34. Baeza JA (2016) Principles of bioprocess control. In: Larroche C, Pandey A, Du G et al (eds) Current developments in biotechnology and bioengineering: bioprocesses, bioreactors and controls. Elsevier Science, Saint Louis, pp 527–561
35. Pörtner R, Platas Barradas O, Frahm B et al (2016) Advanced process and control strategies for bioreactors. In: Larroche C, Pandey A, Du G et al (eds) Current developments in biotechnology and bioengineering: bioprocesses, bioreactors and controls. Elsevier Science, Saint Louis, pp 463–493
36. Fenila F, Shastri Y (2016) Optimal control of enzymatic hydrolysis of lignocellulosic biomass. Resour Effici Technol 2:S96–S104. https://doi.org/10.1016/j.reffit.2016.11.006
37. Moradi H, Saffar-Avval M, Bakhtiari-Nejad F (2011) Nonlinear multivariable control and performance analysis of an air-handling unit. Energ Buildings 43:805–813. https://doi.org/10.1016/j.enbuild.2010.11.022
38. Alford JS (2006) Bioprocess control: advances and challenges. Comput Chem Eng 30:1464–1475. https://doi.org/10.1016/j.compchemeng.2006.05.039
39. Morales-Rodríguez R, Capron M, Hussom JK et al. (2010) Controlled fed-batch operation for improving cellulose hydrolysis in 2G bioethanol production. In: 20th European symposium on computer aided process engineering – ESCAPE20
40. Nyttle VG, Chidambaram M (1993) Fuzzy logic control of a fed-batch fermentor. Bioprocess Eng 9:115–118. https://doi.org/10.1007/BF00369040
41. Álvarez L, García J, Urrego D (2006) Control of a fedbatch bioprocess using nonlinear model predictive control. IFAC Proc 39:347–352. https://doi.org/10.3182/20060402-4-BR-2902.00347
42. Chang L, Liu X, Henson MA (2016) Nonlinear model predictive control of fed-batch fermentations using dynamic flux balance models. J Process Control 42:137–149. https://doi.org/10.1016/j.jprocont.2016.04.012
43. Craven S, Whelan J, Glennon B (2014) Glucose concentration control of a fed-batch mammalian cell bioprocess using a nonlinear model predictive controller. J Process Control 24:344–357. https://doi.org/10.1016/j.jprocont.2014.02.007
44. Li M (2015) Adaptive predictive control by open-loop-feedback-optimal controller for cultivation processes. Dissertation, Jacobs University
45. Frahm B, Lane P, Märkl H et al (2003) Improvement of a mammalian cell culture process by adaptive, model-based dialysis fed-batch cultivation and suppression of apoptosis. Bioprocess Biosyst Eng 26:1–10. https://doi.org/10.1007/s00449-003-0335-z
46. Frahm B, Hass VC, Lane P et al (2003) Fed-Batch-Kultivierung tierischer Zellen - Eine Herausforderung zur adaptiven, modellbasierten Steuerung. Chemi Ingen Tech 75:457–460. https://doi.org/10.1002/cite.200390093
47. Frahm B, Lane P, Atzert H et al (2002) Adaptive, model-based control by the open-loop-feedback-optimal (OLFO) controller for the effective fed-batch cultivation of hybridoma cells. Biotechnol Prog 18:1095–1103. https://doi.org/10.1021/bp020035y
48. Zacher S, Reuter M (2017) Regelungstechnik für Ingenieure. Springer Fachmedien Wiesbaden, Wiesbaden
49. Grüne L, Pannek J (2017) Nonlinear model predictive control. Springer, Cham

50. Hodge DB, Karim MN, Schell DJ et al (2009) Model-based fed-batch for high-solids enzymatic cellulose hydrolysis. Appl Biochem Biotechnol 152:88–107. https://doi.org/10.1007/s12010-008-8217-0
51. Bück A, Casciatori FP, Thoméo JC et al (2015) Model-based control of enzyme yield in solid-state fermentation. Proc Eng 102:362–371. https://doi.org/10.1016/j.proeng.2015.01.163
52. Luttmann R, Munack A, Thoma M (1985) Mathematical modelling, parameter identification and adaptive control of single cell protein processes in tower loop bioreactors. In: Fiechter A, Aiba S, Bungoy HR et al (eds) Agricultural feedstock and waste treatment and engineering, vol 32. Springer, Berlin, pp 95–205
53. Witte VC, Munack A, Märkl H (1996) Mathematische Modellierung und adaptive Prozeßsteuerung der Kultivierung von Cyathus striatus. Zugl.: Hamburg-Harburg, Techn. Univ., Arbeitsbereich Regelungstechnik und Systemdynamik [i.e. Arbeitsbereich Regelungstechnik] und Arbeitsbereich Bioprozess- und Bioverfahrenstechnik, Diss., 1996, Als Ms. gedr. Fortschritt-Berichte/VDI Reihe 17, Biotechnik, vol 144. VDI-Verl., Düsseldorf
54. Patle DS, Ahmad Z, Rangaiah GP (2014) Operator training simulators in the chemical industry: review, issues, and future directions. Rev Chem Eng 30. https://doi.org/10.1515/revce-2013-0027
55. Cameron D, Clausen C, Morton W (2002) Dynamic simulators for operator training. In: Braunschweig B, Gani R (eds) Software architectures and tools for computer aided process engineering, vol 11, 1st edn. Elsevier, Amsterdam, pp 393–431
56. Pörtner R, Platas-Barradas O, Gradkowski J et al (2011) "BioProzessTrainer" as training tool for design of experiments. BMC Proc 5(Suppl 8):P62. https://doi.org/10.1186/1753-6561-5-S8-P62
57. Gerlach I, Hass VC, Brüning S et al (2013) Virtual bioreactor cultivation for operator training and simulation: application to ethanol and protein production. J Chem Technol Biotechnol 88:2159–2168. https://doi.org/10.1002/jctb.4079
58. Reinig G, Winter P, Linge V et al (1998) Training simulators: engineering and use. Chem Eng Technol 21:711–716. https://doi.org/10.1002/(SICI)1521-4125(199809)21:9<711:AID-CEAT711>3.0.CO;2-H
59. Ahmad AL, Low EM, Abd Shukor SR (2010) Safety improvement and operational enhancement via dynamic process simulator: a review. Chem Prod Process Model 5. https://doi.org/10.2202/1934-2659.1502
60. González Hernández Y, Jáuregui Haza UJ, Albasi C et al (2014) Development of a submerged membrane bioreactor simulator: a useful tool for teaching its functioning. Educ Chem Eng 9: e32–e41. https://doi.org/10.1016/j.ece.2014.03.001
61. Gerlach I, Brüning S, Gustavsson R et al (2014) Operator training in recombinant protein production using a structured simulator model. J Biotechnol 177:53–59. https://doi.org/10.1016/j.jbiotec.2014.02.022
62. Ahmad Z, Patle DS, Rangaiah GP (2016) Operator training simulator for biodiesel synthesis from waste cooking oil. Process Saf Environ Prot 99:55–68. https://doi.org/10.1016/j.psep.2015.10.002
63. Balaton MG, Nagy L, Szeifert F (2013) Operator training simulator process model implementation of a batch processing unit in a packaged simulation software. Comput Chem Eng 48:335–344. https://doi.org/10.1016/j.compchemeng.2012.09.005
64. Gerlach I, Mandenius C-F, Hass VC (2015) Operator training simulation for integrating cultivation and homogenisation in protein production. Biotechnol Rep (Amst) 6:91–99. https://doi.org/10.1016/j.btre.2015.03.002
65. Hass VC, Kuhnen F, Schoop K-M (2005) An environment for the development of operator training systems (OTS) from chemical engineering models, vol 20. Elsevier, Amsterdam, pp 289–293. https://doi.org/10.1016/S1570-7946(05)80170-1
66. Brüning S, Gerlach I, Pörtner R et al (2017) Modeling suspension cultures of microbial and mammalian cells with an adaptable six-compartment model. Chem Eng Technol 40:956–966. https://doi.org/10.1002/ceat.201600639

67. R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
68. Xiong Z-Q, Guo M-J, Guo Y-X et al (2010) RQ feedback control for simultaneous improvement of GSH yield and GSH content in *Saccharomyces cerevisiae* T65. Enzym Microb Technol 46:598–602. https://doi.org/10.1016/j.enzmictec.2010.03.003

Adv Biochem Eng Biotechnol (2021) 177: 95–126
https://doi.org/10.1007/10_2020_145

Published online: 11 November 2020

The Kalman Filter for the Supervision of Cultivation Processes

Abdolrahim Yousefi-Darani, Olivier Paquet-Durand, and Bernd Hitzmann

Contents

Abstract In the era of technology and digitalization, the process industries are undergoing a digital transformation. The available process models, advance sensor technologies, enhanced computational power and a broad set of data analytical

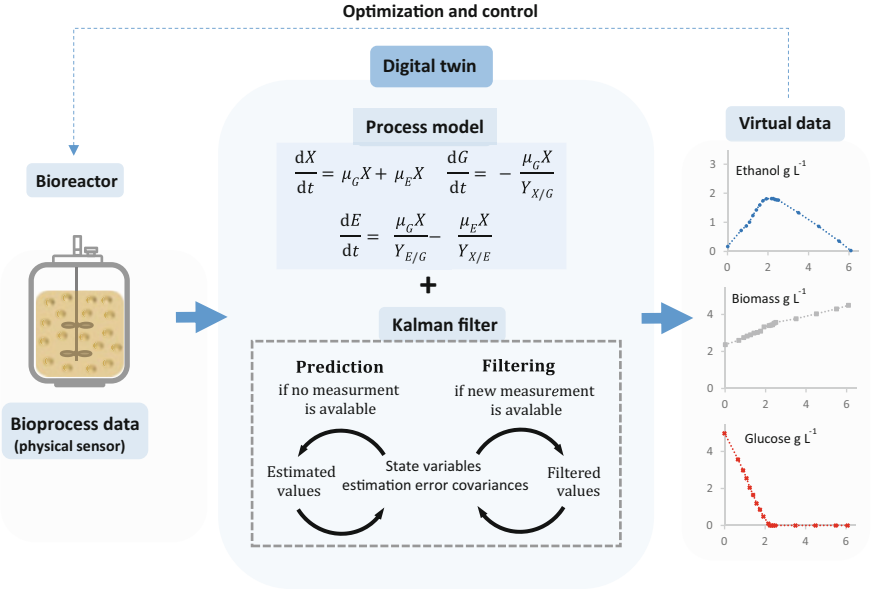
A. Yousefi-Darani (✉), O. Paquet-Durand, and B. Hitzmann
Department of Process Analytics and Cereal Science, Institute of Food Science and Biotechnology, University of Hohenheim, Stuttgart, Germany
e-mail: rahim@uni-hohenheim.de

techniques enable solid bases for digital transformation in the biopharmaceutical industry.

Among various data analytical techniques, the Kalman filter and its non-linear extensions are powerful tools for prediction of reliable process information. The combination of the Kalman filter with a virtual representation of the bioprocess, called digital twin, can provide real-time available process information. Incorporation of such variables in process operation can provide improved control performance with enhanced productivity.

In this chapter the linear discrete Kalman filter, the extended Kalman filter and the unscented Kalman filters are described and a brief overview of applications of the Kalman filter and its non-linear extensions to bioreactors are presented. Furthermore, in a case study an example of the digital twin of the backer's yeast batch cultivation process is presented.

Graphical Abstract A digital twin of a bioreactor mirrors the processes of the real bioreactor. It contains the physical parts, the process model and prediction algorithm to predict the bioprocess variables. These values could be used for optimization and control of the process.



Keywords Bioprocess supervision, Cultivation, Digital twin, Estimation, Kalman filter

Abbreviations

A	State transition matrix
B	Process input transition matrix
C	Measurement model
CKF	Cubature Kalman filter
EKF	Extended Kalman filter
EnKF	Ensemble Kalman filter
F	Jacoby matrix of f()
f()	Non-linear function describing the process change
FIA	Flow injection analysis
H	Jacoby matrix of measurement model
h()	Measurement model
KF	Kalman filter
P	Estimation error covariance matrix
p	Model parameter vector for estimation
Q	Process noise covariance matrix
R	Measurement noise covariance matrix
t	Time
UKF	Unscented Kalman filter
v	Measurement noise vector
w	Process noise vector
x	State variables vector
x(t)	State variable at continuous time k
$\mathbf{x}_{[k]}$	State variable at discrete time k
$\mathbf{x}_{f,[k]}$	Filtered state variable at discrete time k
z	Measurement vector

1 Introduction

Bioprocesses are described as biological systems that are non-linear, complex and unsteady; thus development of precise control systems in order to achieve robust product quality and productivity can be challenging. The control of these processes can be significantly improved by online process monitoring followed by corrective actions. In this context, bioprocess digital twins are helpful tools.

Digital twins are virtual representations of the production process which enable pre-emptive process control by using online data to predict the process outcome in advance. They convert the physical process to a smart process and thus achieve the ultimate goal of the digital transformation. This enables unprecedented possibilities for timely and automated intervention to provide critical decision support during process development [1].

Digital twins mainly consist of a mathematical model which describes the dynamic behaviour observed in a biochemical reactor and a prediction or self-learning algorithm which estimates the cellular component concentrations and the process parameters that cannot be described mechanistically [2, 3].

Bioprocess mathematical models may generally be categorized into algebraic equations and dynamic models. Algebraic equations are developed from mass and component balances, from mass or heat transfer laws or even from elemental balances. Dynamic models usually consist of dynamic balances of conserved quantities in combination with kinetics to describe rate expressions as functions of the state variables. Detailed description of mathematical modelling of bioprocesses is covered by previous authors in greater details than space allows here [4–7]. The goal of this chapter is to highlight state estimation methods with a specific focus on the Kalman filter and its non-linear extensions.

For linear systems, the Luenberger observer and the Kalman filter, whose 60th anniversary occurred in 2020 [8], are the most applied methods for estimating parameters and process variables that cannot be measured directly. In the area of non-linear systems, particle filtering (PF), high gain observers, non-linear extensions of the Kalman filter such as the extended Kalman filter (EKF) and the unscented Kalman filter (UKF) and many others have been proposed. However, due to the simple structure and low computational effort of non-linear extensions of the Kalman filter, these methods have gained more interest, and many research studies have been dedicated to the implementation of such filters for state and parameter estimation in bioprocess technologies. The main objective of this chapter is to discuss the applications of different Kalman filter algorithms in bioprocess technologies. Therefore, this chapter is organized as follows: in the next section, a brief overview of the Kalman filtering theory and its non-linear extensions will be discussed. Applications of the Kalman filter for the supervision of cultivation processes will be given in the third section, followed by a case study evaluating the implementation of an extended Kalman filter for developing a digital twin of the backer's yeast batch cultivation process. In the last section, a conclusion is presented.

2 Kalman Filtering Theory and Its Non-linear Extensions

The Kalman filter is a set of mathematical equations that provides an efficient computational solution of the least-squares method when the considered system is linear and the uncertainties are modelled by Gaussian random variables. When the system state dynamics is non-linear, then certain linearization methods are applied. The most prominent of these algorithms are the extended Kalman filter (EKF) and the unscented Kalman filter (UKF), invented independently by several research groups. Different extensions of the Kalman filters differ in the way the estimation error is calculated. A brief overview of these methods are as follows.

2.1 The Kalman Filter

The Kalman filter is used to provide optimal estimates of unmeasured states for time varying linear systems in the presence of noise by combining information from a process mathematical model with online process measurements. The process model defines the evaluation of the state from time $k-1$ to time k as:

$$\boldsymbol{x}_{[k]} = A\boldsymbol{x}_{[k-1]} + B\boldsymbol{u}_{[k-1]} + \boldsymbol{w}_{[k-1]} \tag{1}$$

where $\boldsymbol{x}$ is the state vector, $\boldsymbol{u}$ is the process input and $\boldsymbol{w}$ is the Gaussian process noise vector that is assumed to be zero-mean with the covariance Q. Matrix A relates the state at the previous time step $k-1$ to the state at the current step k, matrix B relates the control input to the state variables $\boldsymbol{x}$.

The process model is paired with the measurement model that describes the relationship between the state and the measurement at the current time step k as:

$$\boldsymbol{z}_{[k]} = C\boldsymbol{x}_{[k]} + \boldsymbol{v}_{[k]} \tag{2}$$

where $\boldsymbol{z}$ is the measurement vector and $\boldsymbol{v}$ is the Gaussian measurement noise vector which is assumed to be zero-mean with the covariance R. Matrix C relates the state to the measurement $\boldsymbol{z}_{[k]}$. Since the measurements does not exhaustively inform on the current situation of the process, the KF aims to provide an estimate of the process state at time k**,** given the initial state of $\boldsymbol{x_0}$**,** the measurements and the information of the system.

The Kalman filter algorithm consists of two steps which are summarized as follows:

- *Prediction step (time update)***:** Using the initial condition, the process model is used to predict the state variables and the estimation error covariance's until the first measurement is available.

$$\boldsymbol{x}_{[k]} = A\boldsymbol{x}_{[k-1]} + B\boldsymbol{u}_{[k-1]} \tag{3}$$

$$P_{[k]} = AP_{[k-1]}A^T + Q \tag{4}$$

In the above equations, $\boldsymbol{x}_{[k]}$ is the state variables estimate at time k which is deduced from a previous estimation of the state $\boldsymbol{x}_{[k-1]}$ at time $k-1$. The new term $\mathbf{P}$ is called the state error covariance matrix which encrypts the error covariance of the predicted state values. $\mathbf{P}_{[k]}$ is the new prediction error covariance matrix at time k and $\mathbf{P}_{[k-1]}$ is the previous estimated error covariance matrix at time $k-1$. Whenever a measurement is available, a correction step is performed:

- *Correction step (measurement update)***:** In this step the predicted model estimates are combined with the measured values to provide corrected estimates.

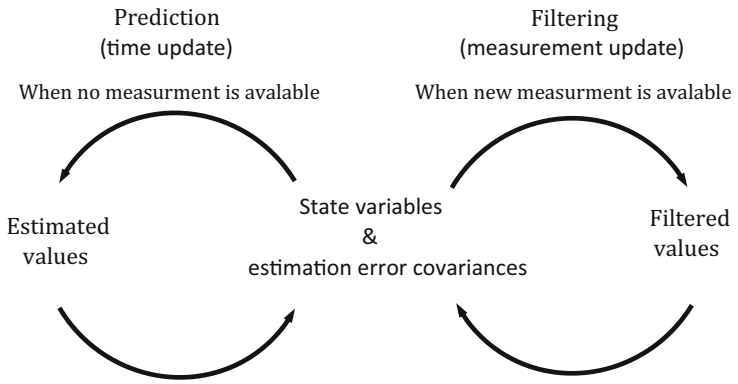


Fig. 1 The flow chart of the Kalman filter algorithm

$$\boldsymbol{x}_{f,[k]} = \boldsymbol{x}_{[k]} + K_{[k]}\left(\boldsymbol{z}_{[k]} - C\boldsymbol{x}_{[k]}\right) \tag{5}$$

$$P_{f,[k]} = P_{[k]}\left(1 - K_{[k]}C\right)^2 + K^2R \tag{6}$$

$$K_{[k]} = P_{[k]}C^T\left(R + CP_{[k]}C^T\right)^{-1} \tag{7}$$

The measurement prediction error, reflects the discrepancy between the true measurements $\boldsymbol{z}_{[k]}$ and the predicted measurements $C\boldsymbol{x}_{[k]}$. The difference of both is multiplied by the so called Kalman gain and used to update the estimated state variables. Therefore the filtered state variables $\boldsymbol{x}_{f,\,[k]}$ are obtained. In the similar manner, the filtered estimation error covariance $P_{f,\,[k]}$ is obtained. $K_{[k]}$ is chosen to minimizes the estimated error covariance

$$\frac{dP_f}{dK} = 0 \tag{8}$$

The measurement error variance must be compared with the estimation error variance to see how the filter is acting. For this purpose, a very rough treatment is necessary:

If $R \ll C\mathrm{P}_{[k]}C^T$ then $K \approx C^{-1}$ and $\boldsymbol{x}_{f,\,[k]} \approx C^{-1}\boldsymbol{z}_{[k]}$; so the filtered is almost determined by the measured.

If $R \gg C\mathrm{P}_{[k]}C^T$ then $\boldsymbol{x}_{f,\,[k]} \approx \boldsymbol{x}_{[k]}$; the filtered value is almost the estimated one and no influence of the measurement will be obtained.

With the filtered values as initial condition the simulation of the process as well as the estimation error covariance's can be carried out until the next measurement is obtained and everything repeats again. The flow chart of the Kalman filter algorithm is presented in Fig. 1.

2.2 Continuous-Discrete Extended Kalman Filter

As described in the previous section, the Kalman filter addresses the general problem of trying to estimate the state of a process that is governed by a linear differential equation system. In non-linear dynamic systems, the process model or the measurement model cannot be determined with multiplication of vectors and matrices. For such systems, a linearization should be performed. The linearization can be performed by different methods. The essential difference among different versions of the Kalman filters (extended Kalman filter, unscented Kalman filter and ensemble Kalman filter) consists in how they calculate the estimation error. A Kalman filter that linearizes about the current mean and covariance is referred to as an extended Kalman filter (EKF). A non-linear dynamic system can be described by the following differential equation:

$$\frac{dx(t)}{dt} = f(x(t), u(t)) + w(t) \tag{9}$$

With discrete measurements that are:

$$\boldsymbol{z}_{[k]} = h\left[\boldsymbol{x}\left(\boldsymbol{t}_{[k]}\right)\right] + \boldsymbol{v}_{[k]} \tag{10}$$

The differential equation provide the continuous part, the measurements are the discrete part, where f is a non-linear function of the state variables $\boldsymbol{x}$ and the control input $\boldsymbol{u}$. The non-linear function h in the measurement equation relates the current state to the measurement $\boldsymbol{z}_{[k]}$. $\boldsymbol{w}$ and $\boldsymbol{v}$ are, respectively, the process noise vector and the measurement noise vector. These noises are assumed to be zero mean, white, and independent of each other, with respective covariance matrices Q and R.

To calculate the estimation error covariance matrix, the following differential equations have to be solved in parallel to the state differential equation.

$$\frac{d\boldsymbol{P}(\boldsymbol{t})}{dt} = F(t)P(t) + P(t)F^T(t) + Q \tag{11}$$

Here the Jacobian matrix is used, which is given by the following equation:

$$F = \left.\frac{\partial f}{\partial x}\right|_{x(t),\ u(t)} \tag{12}$$

The filtering is performed as follows:

$$K_{[k]} = P(t_k)H^T(t_k)\left[H(t_k)P(t_k)H^T(t_k) + R\right]^{-1} \tag{13}$$

$$\boldsymbol{x}_f(t_{[k]}) = \boldsymbol{x}(t_{[k]}) + K_{[k]}\left[\boldsymbol{z}_{[k]} - h\left[\boldsymbol{x}(\boldsymbol{t}_{[k]})\right]\right] \tag{14}$$

$$P_f(t_{[k]}) = \left[I - K_{[k]}H_{[k]}\right]P(t_{[k]})\left[I - K_{[k]}H_{[k]}\right]^T + K_{[k]}RK_{[k]}^T \tag{15}$$

where $H_{[k]}$ is the Jacoby matrix of h[]:

$$H_{[k]} = \left.\frac{\partial h}{\partial x}\right|_{x_{[k]}} \tag{16}$$

Correspondingly to the KF algorithm, the EKF algorithm consists of two main parts including prediction step and the correction step.

As mentioned above, the basic framework for the EKF involves state estimation of a non-linear dynamic system. However, in some cases, prediction of $\boldsymbol{x_k}$ requires coupling both state estimation and parameter estimation [9]. Here a process model parameter $p(t)$ is considered to be time dependent and can be estimated by adding the parameter as an additional state variable whose differential equation is then given as

$$\frac{d\boldsymbol{p}(t)}{dt} = 0 \tag{17}$$

At every time step, the current estimate of the parameter $p(t)$ is used in the measurement filter. In the joint estimation method, model state variables and model parameters are included in a single joint state vector. Parameter estimation evolves in time along with state estimation, as observations are assimilated [10].

Other alternatives for parameter estimation with the KF include calibrating parameters outside the KF calculation with an outer optimisation routine [11–13], and parameter estimation in steady-state KF calculations where observations are climatological averages over the entire time period of interest [14], but in both of these two approaches the parameter estimation part of the calculation considers all observations at once rather than sequentially.

2.3 Other Non-linear Extensions of the Kalman Filter

As mentioned previously, when the system is non-linear and can be well approximated by linearization, then the EKF is a good option for state estimation; however EKF is not optimal if the system is highly non-linear, this is because only the mean is propagated through the non-linearity [15]. The unscented Kalman filter (UKF) is another non-linear extension of the Kalman filter which is a discrete time filtering algorithm. The UKF utilizes the unscented transformation for computing approximate solutions to the filtering problems.

A general framework for state estimation based on the UKF for this state space model is presented as follows:

In the first step, the initial values for the state and covariance estimation have to be set. Following this, the recursive estimation is performed by the prediction and correction steps. Within the prediction step, a priori state and covariance estimation utilizing the process model is performed. Using the unscented transformation, a set of sigma points are chosen. These sigma points characterize the current probability density function. Each point from the sigma matrix is propagated through the process model to calculate the estimations of state variables and the error covariance. Following this, a correction step is preformed when a measurement is received. This leads to the estimations of the filtered state variables and the filtered error covariance by calculating the Kalman gain.

The UKF has been used in various fields for non-linear sate estimations. However a couple of alternative approaches have emerged over the last few years, namely, the ensemble Kalman filter (EnKF) and the cubature Kalman filter (CKF) which are widely used when the process model is of extremely high order and non-linear, the initial states are highly uncertain and a large number of measurements are available [16, 17].

Similar to the UKF, the EnKF and CKF select a set of sample points (sigma points) in order to deal with the non-linearity of the system. In high-dimension systems, the weights of the sigma points in the UKF are prone to be negative, leading to low estimation accuracy.

In EnKF the error covariances are estimated approximately using an ensemble of model forecasts. The main concept behind the formulation of the EnKF is that if the dynamical model is expressed as a stochastic differential equation, the prediction error statistics, which are described by the Fokker–Plank equation, can be estimated using ensemble integrations, and the error covariance matrices can be calculated by integrating the ensemble of model states [16].

The cubature Kalman filter uses the spherical–radial cubature rule to generate some weighted sampling points to approximate integral in Bayesian estimation. A brief overview of the unscented Kalman filtering and sigma point filtering in general are given by van der Merwe [18].

3 Application of Kalman Filters in Bioprocess Monitoring

Here 41 recent published articles [19–60] in the period of 1991–2020 on application of the Kalman filter and its extensions for state and parameter estimation in bioprocesses are discussed. Due to space limitation, only some of the reported articles are presented in Table 1. The table is organized by classifying the articles into different categories, which include the type of the Kalman filter and the applied process model, the type of microorganism and the cultivation process mode, the measured process variable(s) and the objective of the filtering algorithm. This table would help understanding how the Kalman filter was explored chronologically to date. It should be mentioned that in some works more than one Kalman filter

Table 1 Extended Kalman filter application for cultivation processes

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
Extended Kalman filter/experimental application	Batch cultivation/*E. coli*	Dissolved oxygen mass balance	Noise filtering from dissolved oxygen measurements	Dissolved oxygen	Lee et al. [19]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Material balance equation with Monod growth rate kinetics	Parameter estimation and substrate prediction	Glucose concentration with FIA	Hitzman et al. [32]
Kalman filter/experimental application	Batch cultivation/ *S. cerevisiae*	Ideal stirred tank reactor model with Monod growth kinetics (glucose and ethanol as limiting substrates)	Noise filtering from predicted bioprocess variables	Biomass, glucose, and ethanol (with ultrasonic velocity)	Cha and Hitzmann [36]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	A model for an ideal stirred tank reactor in combination with Monod growth kinetics	Noise filtering from predicted glucose	Glucose concentration with flow injection analyses (FIA)	Arndt and Hitzmann [37]
Extended Kalman filter/ simulation	Fed-batch cultivation/ *S. cerevisiae*	Cybernetic model of Jones and Kompala	Filtering out noise from the feed stream	Dilution rate or the gas–liquid mass transfer coefficient for oxygen	Patnaik [39]
Extended Kalman filter/ simulation	Fed-batch cultivation/*E. coli*	General dynamic model of bioreactors with Monod growth kinetics	Parameter estimation and biomass prediction	Dissolved and exhaust oxygen and carbon dioxide	Rocha et al. [40]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *Bordetella pertussis*	A model with two parameters which are calculated using separate experiments	Estimation of specific growth rate, biomass, and oxygen mass transfer	Dissolved oxygen	Soons et al. [42]

(continued)

Table 1 (continued)

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
Unscented Kalman filter/experimental application	Fed-batch cultivation/ hybridoma cell culture	Overflow metabolism model	Noise reduction from predicted values	Predicted specific uptake and production rate	Henry et al. [41]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Ideal stirred tank reactor model with Monod growth kinetics	Parameter, biomass, and glucose prediction	Glucose concentration with FIA	Klockow et al. [43]
Extended Kalman filter/experimental application	Fed-batch cultivation/*E. coli*	General dynamic model of bioreactors with Monod growth kinetics	Estimation of biomass, glucose, and acetate	Dissolved oxygen and carbon dioxide	Veloso et al. [44]
Unscented Kalman filter/ simulation	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass in the headspace with Monod growth kinetics	Estimation of biomass and substrate concentrations	Dissolved oxygen and carbon dioxide	Jianlin et al. [46]
Unscented Kalman filter/ simulation	Fed-batch/ hybridoma cell	Material balance equation with Monod growth kinetic	Prediction of acetate and glucose concentration	Biomass and dissolved oxygen	Dewasme et al. [48]
Extended Kalman filter/ simulation	Batch cultivation/ *S. cerevisiae*	Unstructured model for alcoholic fermentation with immobilized cells using Monod growth kinetics	Estimation of product, substrate, and biomass concentrations	Glucose and ethanol	Popova et al. [49]
Extended Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass in the headspace with Monod	Estimation of substrate and biomass concentrations	Substrate and biomass concentration with NIR spectrometer	Krämer and King [54]

(continued)

Table 1 (continued)

Estimator/ application type	Cultivation type/ microorganism	Process model	Objective	Measured state	Reference
		growth kinetics			
Unscented Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass with Monod growth kinetics	Biomass and specific biomass growth rat estimation	Oxygen uptake and CO_2formation rate	Simutis and Lübert [55]
Sigma point Kalman filter/experimental application	Fed-batch cultivation/ *S. cerevisiae*	Mass balance of substrate and biomass in the headspace with Monod growth kinetics	Estimation of substrate and biomass concentrations	Substrate and biomass concentration with NIR spectrometer	Krämer and King [57]
Extended Kalman filter/ simulation	Fed-batch cultivation/ *S. cerevisiae*	Material balance equation with Monod growth rate kinetics	Ethanol prediction and state estimation	Temperature, do and substrate concentration	Lisci and Tronci et al. [60]

algorithm are examined. More detailed description of each category for all publications is presented in the following part of this section.

3.1 Type of Kalman Filter

According to the type of Kalman filter algorithm, the literature presented indicates there exist a considerable number of articles on implementation of EKF for state and parameter estimation. More than 60% of the applications (28 articles) have implemented EKF algorithms for their process. This is due to the fact that the cultivation process of microorganisms is a complex non-linear biochemical process and the EKF is a well-known state estimation method for non-linear systems. The linear Kalman filter which is almost exclusively used for state estimation in linear systems have also been used by some authors (3 articles). Although the EKF shows good prediction results and is widely used in literature, it presents some disadvantages. It is reliable for systems which are almost linear on the time scale of the update intervals; it requires the calculation of Jacobians at each time step, which may be difficult to obtain for higher order systems; it does linear approximations of the system at a given time instant, which may introduce errors in the estimation, leading then the state to diverge over time [9, 15]. For instance, in continuous or fed-batch

cultivations, despite continuous supply by a feed, the substrate concentration can drop to zero as the cell takes it up very fast. In such cultivations, linearization in the time and measurement update can lead to significant inaccuracies in the process, while the EKF assumes a certain probability for substrate concentrations below zero, even though this is physically impossible [54]. Therefore in recent years, application of other non-linear extensions of the Kalman filter is used. For example, Fernandes et al. [54] have implemented an UKF algorithm in order to estimate glucose and glutamine from biomass, lactate and ammonia measurement during fed-batch cultivation of hybridoma cells. The predictions were compared to the ones obtained with an EKF; they have reported the UKF achieves better level of accuracy. Krämer and King [57] have implemented a UKF in fed-batch cultivation of *S. cerevisiae* for noise filtering from predicted biomass values with NIR spectrometer. In another study, the same authors [54] have implemented an EKF for the same process. The authors have reported accurate predicted values in both studies; however there is no comparison between the two methods. Other types of the non-linear Kalman filtering method have also been reported in literature. Zhao et al. [53] have implemented a CKF for incorporating delayed measurements of biomass, substrate, and product concentration in fed-batch cultivation for penicillin production. Bavdekar et al. [47] have implemented an EnKF for overcoming delayed measurements of biomass, substrate and ethanol concentration in fed-batch cultivation of *S. cerevisiae*. Addressing the same delay problem Klockow et al. [43] complemented a ring buffer by an EKF and got satisfied results.

In order to indicate which Kalman filter extension describes the process better, numerical simulation runs are required. According to this perspective, a closer look to the presented articles indicates that most studies (31 articles) had relied on practical applications and simulation studies have been reported only 12 times.

3.2 Microorganism

Regarding the type of microorganism, the articles show that the majority of the research has focused on applying the Kalman filter or its extensions for state or parameter estimation during the cultivation of *S. cerevisiae* (19 articles) and *E. coli* (7 articles). The importance of these microorganisms for the biopharmaceutical industry is widely recognized, as *E. coli* and *S. cerevisiae* are the most important host microorganism used to produce recombinant proteins [58]. In addition, *S. cerevisiae* is also widely used for the production of the backers yeast as well as wine and beer. Only a few articles demonstrate state estimation in the cultivation process of other microorganisms. For instance, some authors have implemented state estimation methods for prediction of substrate and product concentration during cultivations of *Candida utilis* [30], *Penicillium chrysogenum* [46, 53] and *Kluyveromyces marxianus* [34].

3.3 Cultivation Mode

From an operational point of view, cultivation of microorganisms can be performed in batch, fed-batch and continuous modes. In fed-batch cultivation modes, set point control of the substrate concentration by manipulating the input flow rate is a matter of particular economic and scientific interest. In order to have an efficient control system, sufficient knowledge about the process state variables is required, which can be achieved by the state estimation methods such as the Kalman filter or its extensions. Therefore, previous studies have almost exclusively focused on the application of state estimation methods for fed-batch cultivations (34 publications). However, online monitoring and estimation of state variables in batch cultivations is also crucial in order to monitor the state and if necessary may improve it to achieve high productivity over the process. For instance, controlling the level of dissolved oxygen (DO) in the fermentation broth, effects the rate of microbial metabolism. Accordingly, Lee et al. [19] have implemented an EKF for noise filtering of dissolved oxygen measurements which were used for controlling the DO levels in batch cultivation of *E. coli*. This approach and, more generally, online monitoring and state estimation of variables in batch cultivations remain briefly addressed in the literature.

3.4 Bioprocess Phase

Mixing of medium and pre-cultures are performed during upstream processing phase and separation and purification of the product from biomass is performed during the downstream processing phase. In order to optimize cell growth and maximize the product yield, online monitoring and a tight control is required during both phases. The presented articles show there have been numerous studies to investigate the application of state estimation methods during the cultivation phase (39 papers). However, the articles indicate that only two authors had examined the application of Kalman filtering methods for state and variable estimation in downstream processing. For efficient and robust process development in the downstream processing phase, knowledge of the location and concentration of the product and key contaminants is also crucial. Holwill et al. [28] have used a low technology detection system involving the measurement of rate of change of absorbance at a single wavelength after addition of reagent to a representative sample stream. This provided online data detailing the performance of a continuous precipitation process. This information as well as a mathematical model which describes the fractional protein perception were fed into a control algorithm which was programmed to maintain predefined set points by feedback control through adjustments to the overall feed saturation. The Kalman filter was used for estimating the parameters of the model. Feidl et al. [59] developed a state estimation procedure for estimation

of antibody concentration by combining information coming from kinetic model and a Raman analyser, in the frame of an extended Kalman filter approach (EKF).

3.5 Measurement Device

An overview of measurement devices that are appropriate for the operation of bioprocesses is presented by Sonnleitner [61]. More specific details of different types of sensors and their measurement principles can be found in literature [62, 63]. The literature presented indicate that in *E. coli* cultivation, most authors have employed DO and CO_2 measurements from the exit gas or glucose measurements using flow injection analysis as the measurement in the Kalman filter algorithm. On the other hand, in *S. cerevisiae* cultivations, besides DO, CO_2 and glucose measurements, biomass measurements have also been widely applied. For example, Dewasme et al. [48] applied biomass measurements for their KF during an *E. coli* cultivation.

3.6 Process Model

According to the articles presented, the general mass balance equations are the most common mathematical approach used for describing the process in state observing algorithms. An overview of typical models applied to bioprocesses is presented by Chhatre [64]. A wide variety of growth kinetics are developed for modelling of particular bioprocesses. The Monod growth model [65] is the most applied method for calculating the growth kinetics of microorganisms; it corresponds to a rational function in which the specific growth rate μ is only a function of a single limiting substrate concentration and is subjected to substrate saturation when $S \gg K_s$.

$$\mu = \mu_{\max} \frac{S}{K_s + S} \tag{18}$$

where $\mu_{\max}$ is the maximum specific growth rate, K_s is the Monod half-saturation constant, and S is the concentration of the limiting substrate. In the mentioned articles, all of the authors, which were growing *S. cerevisiae* and *E. coli,* have implemented the Monod growth kinetics. A modified Monod model was applied by Patnaik [35, 38] which is described in detail by Henson and Seborg [66] or Jones and Kompala [67]. Application of other methods for calculating the growth kinetics such as the Contois growth model [68] has also been reported. A feature of the Contois growth model is that growth rate depends upon the concentrations of both substrate and cell mass with the consequence that an inhibition is present at high cell concentrations. This growth kinetic has been implemented in a process model describing the growth behaviour of *Penicillium chrysogenum* in fed-batch

cultivations. A modified Contois model was applied by Jianlin et al. [48] and Zhao et al. [53] in an UKF and CKF algorithm for biomass and substrate prediction, respectively. The growth rate can also be represented by artificial neural networks. However this kind of models is not applied often in combination with a KF. Zorzetto and Wilson [27] have applied a hybrid model in an EKF algorithm which is based on the theory of limited respiratory with using artificial neural network for predicting the growth rates during fed-batch cultivation of *S. cerevisiae.*

Most of the process models which are reported in literature and are used in the Kalman filter algorithms are considered to be ideal stirred tank reactors, whereas production-scale operations are corrupted by noise. This problem is more sever in large-scale operations than in laboratory-scale fermentations [35]. This can describe why all applications of state estimation methods presented in Table 1 are performed in laboratory-scale bioreactors (most cultivations are performed in a 2–5 L bioreactor and one cultivation [57] have been performed in a 22 L bioreactor).

4 An Extended Kalman Filter for the Monitoring of a Yeast Cultivation

The integration of gas sensor array data in a non-linear state estimator has not been discussed previously in the literature. Yousefi-Darani et al. [69] have designed and implemented a model-based calibrated gas sensor array for online measurement of ethanol concentration in batch cultivation with the yeast *S. cerevisiae*. However the predicted values are only available every 5 min. Therefore in this work, in order to have continues values of ethanol concentration as well as the values of biomass, glucose and the maximal growth rates, we have implemented an EKF. In addition, the whole estimation producer could be considered as a digital twin of the baker's yeast batch cultivation process, which could be used for process optimization and control.

4.1 The Cultivation Process

The cultivation of *Saccharomyces cerevisiae* (fresh baker's yeast, Oma's Ur-Hefe) was carried out in a 2.5 L bioreactor (Minifors, Infors HT, Bottmingen, Switzerland) with a vessel of stainless steel working volume of 1.35 L equipped with a temperature (set point of 30°C) and pH (set point pH = 5) control unit. The aeration and agitation rates were kept constant at 3.5 L min^{-1} and 500 rpm, respectively. For the pre-culture, 5 g of the baker's yeast was suspended into 100 mL medium containing 0.34 g L^{-1} $MgSO_4 \cdot 7H_2O$, 0.42 g L^{-1} $CaCl_2 \cdot 2H_2O$, 4.5 g L^{-1} $(NH_4)_2SO_4$, 1.9 g L^{-1} $(NH_4)_2HPO_4$, 0.9 g L^{-1} KCl. The inoculation was performed after 10 min of shaking. The same medium supplemented with glucose to a final concentration of

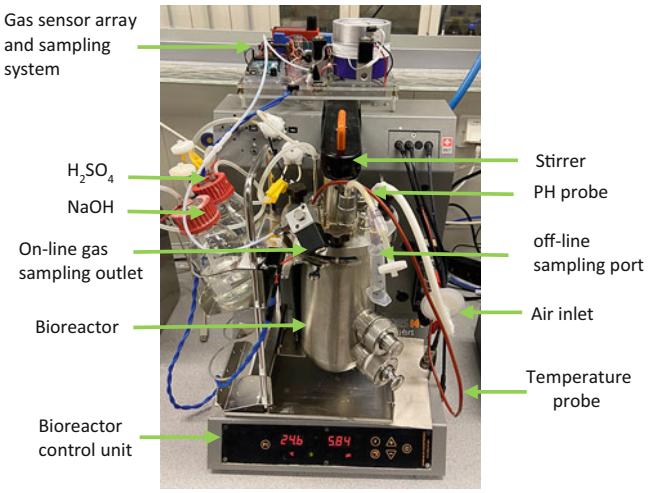


Fig. 2 Overview of the experimental setup

5 g L^{-1} as well as 1 mL L^{-1} trace elements solution (0.015 g L^{-1} $FeCl_3 \cdot 6H_2O$, 9 mg L^{-1} $ZnSO_4 \cdot 7H_2O$, 10.5 mg L^{-1} $MnSO_4 \cdot 2H_2O$, and 2.4 mg L^{-1} $CuSO_4$ $5H_2O$) and 1 mL L^{-1} vitamin solution (0.06 g L^{-1} myoinositol, 0.03 g L^{-1} Ca-pantothenate, 6 mg L^{-1} thiamine HCl, 1.5 mg L^{-1} pyridoxine HCl, and 0.03 mg L^{-1} biotin) was used for the cultivation. The experimental setup is presented in Fig. 2.

4.2 EKF Algorithm

The EKF uses discrete measurements of ethanol from the gas sensor array and estimates continuous online values of ethanol, biomass and glucose concentrations as well as the maximal growth rates in *S. cerevisiae* batch cultivation. A detailed description of the working principle of the EKF is presented in Sect. 2.2.

The EKF was implemented using the software Matlab® 2019a (version 9.6.0); the "Symbolic Math" toolbox (version 8.3) was used to calculate the estimation error covariance differential equation matrix (25 equations). For all calculations, a normal office PC (Intel Core® i5 8,500 with 8 GiB of RAM) with Window 10 was used. For the simulation, the system of in total 30 (5 + 25) differential equations was solved numerically using the explicit, Runge–Kutta-based ode45 method from Matlab. The Matlab code can be found in the appendix.

4.3 Online Ethanol Measurements

The online ethanol measurements were performed in a self-developed system equipped with commercially available metal oxide semiconductor (MOS) gas sensors (TGS 822, TGS 813 and MQ3). The sensors were located in a measuring chamber with a volume of 250 mL and operated in two cycles: a measurement cycle and a washing cycle. During the measurement cycle, the headspace gas was pumped into the measurement chamber for 10 s at a flow rate of 400 mL min^{-1} with a diaphragm pump (Schwarzer Precision, Essen, Germany). Then the chamber was flushed by pure oxygen for regeneration. A peak-shaped measurement signal is obtained, which was evaluated by using a chemometric model, which is described in detail in the literature [69]. Therefore, every 5 min a new ethanol measurement value is used by the Kalman filter. Figure 3 presents a schematic diagram of the online ethanol measurement system and the EKF for continuous state variables and parameter estimation.

Note that the EKF was carried out after the experiments were performed. The results, however, carry over to a true online application where the data is not analysed or modified in retrospect.

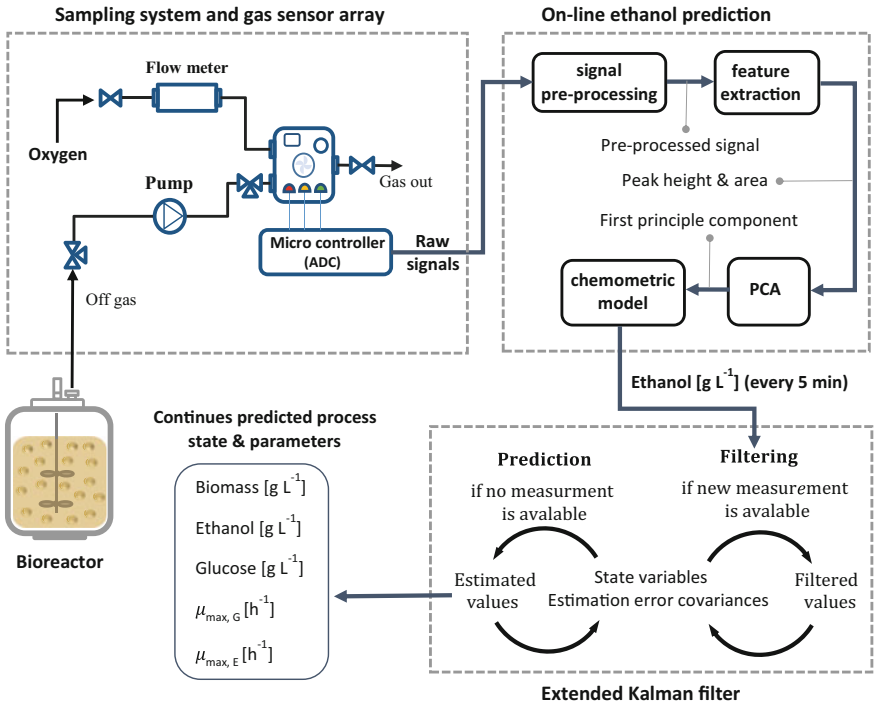


Fig. 3 Schematic diagram of the online ethanol measurement system and the EKF for continuous state variables and parameter estimation

4.4 Offline Measurements

For offline analysis, samples were regularly taken from the bioreactor and placed in pre-weighed and pre-dried micro centrifuge tubes. For biomass determination, the sample without supernatant were dried for 24 h at 103°C and after cooling for 30 min weighed. Using the filtrated supernatant (pore size filter, 0.45 µm, polypropylene membrane, VWR, Darmstadt, Germany), glucose and ethanol were determined by HPLC (ProStar, Variant, Walnut Creek, CA, USA); injection of 20 µL into a Rezex ROA-organic acid H+ (8%) column (Phenomenex, Aschaffenburg, Germany) and operated at 70°C with 5 mM H_2SO_4 as an eluent at 0.6 mL min^{-1} flow rate; software GalaxieTM Chromatography (Varian, Walnut Creek, CA, USA). The offline values were not used during the estimation of the state variables and are only taken to show that the estimates are accurate.

4.5 State Equations of the Cultivation Process

As bioreactor an ideal stirred tank reactor was assumed. As state variables, the biomass, glucose and ethanol concentrations as well as the maximal specific growth rate on glucose and ethanol were applied. Therefore, the following state equations are obtained:

$$\frac{d}{dt}\begin{bmatrix} X \\ G \\ E \\ \mu_{max,G} \\ \mu_{max,E} \end{bmatrix} = \begin{bmatrix} (\mu_G + \mu_E)X \\ -\frac{\mu_G}{Y_{GX}}X \\ \left(\frac{\mu_G}{Y_{GE}} - \frac{\mu_E}{Y_{EX}}\right)X \\ 0 \\ 0 \end{bmatrix} \tag{19}$$

were μ_G and μ_E are given as

$$\mu_G = \frac{\mu_{max,G} \cdot G}{K_G + G} \tag{20}$$

$$\mu_E = \frac{\mu_{max,E} \cdot E}{K_E + E} \cdot \left(1 - \frac{\mu_G}{\mu_{max,G}}\right)^2 \tag{21}$$

As one can see from the state equation, the Kalman filter is used to estimate the maximum specific growth rate on glucose $\mu_{max,\ G}$ and on ethanol $\mu_{max,\ E}$. The importance of the specific growth rate for the assessment of a cultivation is discussed by Galvanauskas et al. [70].

Table 2 Parameter values used for the simulation model

Parameter	Value	Description
K_G	0.1 gL^{-1}	Monod constant glucose
K_E	0.1 gL^{-1}	Monod constant ethanol
Y_{GX}	0.17 gg^{-1}	Conversion factor glucose to biomass
Y_{GE}	0.46 gg^{-1}	Conversion factor glucose to ethanol
Y_{EX}	0.6 gg^{-1}	Conversion factor ethanol to biomass

Table 3 Initial conditions for the extended Kalman filter

Parameter	Value	Description
$X_{t=0}$	2.4 gL^{-1}	Initial biomass concentration
$G_{t=0}$	5.0 gL^{-1}	Initial glucose concentration
$E_{t=0}$	0.1 gL^{-1}	Initial ethanol concentration
$\mu_{max,\,G}$	0.14 h^{-1}	Initial maximal growth rate on glucose
$\mu_{max,\,E}$	0.07 h^{-1}	Initial maximal growth rate on ethanol
$P_{t=0}$	$\begin{pmatrix} 0.02\ g^2L^2 & 0 & 0 & 0 & 0 \\ 0 & 0.02\ g^2L^2 & 0 & 0 & 0 \\ 0 & 0 & 0.02\ g^2L^{-2} & 0 & 0 \\ 0 & 0 & 0 & 0.02\ h^{-2} & 0 \\ 0 & 0 & 0 & 0 & 0.02\ h^{-2} \end{pmatrix}$	Initial estimation error covariance matrix

The extension to the ordinary Monod model for μ_E is applied, so that the transformation from glucose consumption to ethanol consumption is modelled. In Tables 2, 3, and 4 the parameters of the model as well as the initial values for the state equations and the initial values of the estimation error covariance are presented.

The Matlab code as well as the measured off- and online data of this example can be found in the appendix.

4.6 Results

In Fig. 4 the online and offline measured values of ethanol, the offline measured values of biomass and glucose as well as all the Kalman filter estimated values of all three bioprocess variables can be seen.

Figure 4 indicates the typical diauxic growth pattern of baker's yeast on glucose is obtained. First the glucose is consumed and biomass and ethanol are produced,

Table 4 Estimated measurement noise and process noise as well as measurement model for the EKF

Parameter	Value	Description
R	$0.0225\ g^2L^{-2}$	Measurement noise variance
Q	$\begin{pmatrix} 0.001\ g^2L^2h^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0.001\ g^2L^2h^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0.001\ g^2L^{-2}h^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0.005\ h^{-3} & 0 \\ 0 & 0 & 0 & 0 & 0.005\ h^{-3} \end{pmatrix}$	Process noise covariance matrix
z	$(0\ 0\ 1\ 0\ 0)$	Measurement matrix (just ethanol is measured)

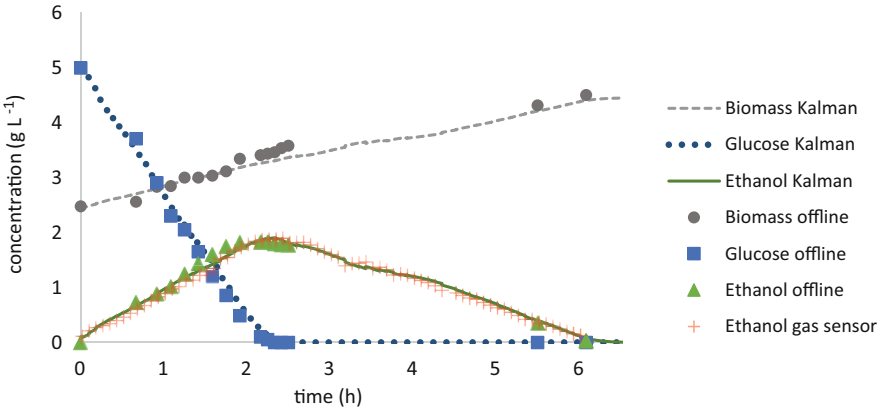


Fig. 4 Online and offline values for biomass, glucose and ethanol as well as EKF estimates for these values

Table 5 Prediction error of EKF values compared to offline measurements

	Glucose	Ethanol	Biomass
RMSEP	0.12 g L^{-1}	0.14 g L^{-1}	0.12 g L^{-1}
Error	5.6%	2.8%	6.2%
R^2	0.96	0.99	0.97

then ethanol is converted to biomass. The offline measurements and its corresponding estimated values fit quite well together as can be seen in Table 5.

The root mean squared error of prediction (RMSEP) of glucose is 0.12 g L^{-1}. The ethanol offline values during glucose consumption are mostly higher than the online measured and the predicted ones; in overall their RMSEP is 0.14 g L^{-1}. All ethanol online measurements seems to be a little bit shifted in time compared to the offline values, which might indicate the time delay due to gas transport from the fermentation broth through the headspace of the reactor to the measurement system. The biomass has a RMSEP of 0.12 g L^{-1}, but the highest deviation can be seen shortly after ethanol is used as substrate. The values shortly before ethanol consumption might not be predicted accurately, because the model describing the switching from glucose to ethanol might be suboptimal.

In order to investigate the influence of the measurement frequency on the performance of the EKF, we decreased the measurement frequency of the online ethanol measurements to one per hour. The results of the estimated values with the EKF are presented in Fig. 5.

Still the overall behaviour of the estimated values is the same. However, the sampling frequency has an influence on the corrections of the estimated state during filtering. Larger step changes are observed in the estimated values whenever a new measurement is available. However, even if the sampling frequency is changed to one per hour, the overall behaviour is predicted well.

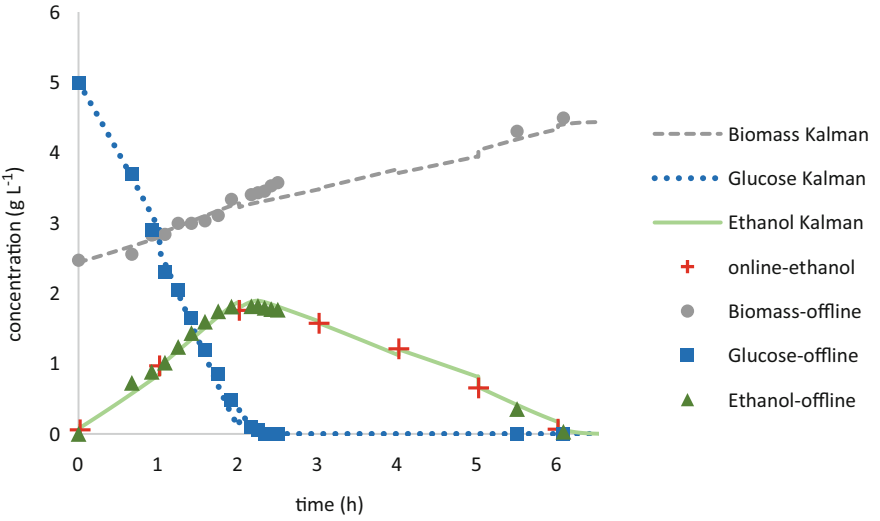


Fig. 5 Online (every 1 h) and offline values for biomass, glucose and ethanol as well as EKF estimates for these values

Obviously with a higher sampling frequency, these step changes are smaller. Nevertheless, with a 5 min sampling time, the EKF was able to follow the true states of the system with a reasonably small error. More detailed information about the influence of the sampling frequency on the accuracy of the Kalman filter estimates can be found in literature [71, 72].

The EKF was also used for predicting the specific growth rates and their maximum values.

In Fig. 6 the estimated maximum specific growth rates with respect to glucose $\mu_{max,\ G}$ and ethanol $\mu_{max,\ E}$ as well as specific growth rates itself (μ_G and μ_E for glucose and ethanol respectively) are presented.

After inoculation, the specific growth rate and its maximum value with respect to glucose are increasing from 0.14 h^{-1} to more than 0.18 h^{-1}. However shortly thereafter they decrease again. This indicates the high sensitivity of the estimation values due to the measurement noise variance R and the process noise variance with respect to $\mu_{max,\ G}$, which is Q [4]. The smaller the R and the higher the Q [4], the more the estimated values will rely upon the measurements and as a consequence the filtered values might be changed, if the measured and estimated values deviate from each other. The more glucose is consumed, the larger will be the difference of $\mu_{max,\ G}$ and μ_G, due to the Monod growth kinetics. If the glucose is almost depleted, the extension to the Monod model on ethanol contributes to increasing growth on ethanol. Shortly after 2 h cultivation time, the transition from glucose to ethanol as substrate takes place. The maximum specific growth rate on ethanol $\mu_{max,\ E}$, which has not changed during the growth on glucose starts to increase. According to the typical Monod behaviour, before ethanol is depleted, due to the low substrate concentration, $\mu_{max,E}$ should be almost constant while μ_E should be increasing.

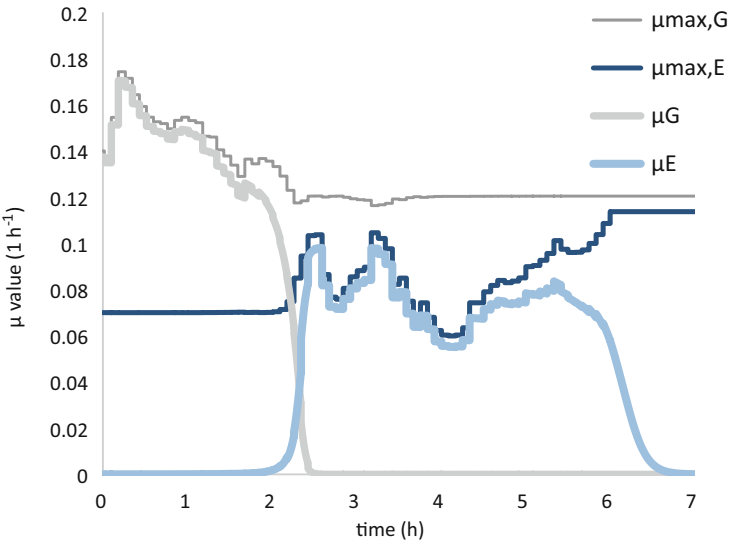


Fig. 6 Estimated maximum specific growth rates with respect to glucose $\mu_{max,\,G}$ and ethanol $\mu_{max,\,E}$ as well as the specific growth rates (μ_G and μ_E for glucose and ethanol, respectively)

However this is not observed in Fig. 6 which is due to the fluctuation of the measured and estimated ethanol concentration.

5 Conclusion

In this chapter, the working principles as well as an overview of Kalman filter applications for state and parameter estimation in bioprocesses has been presented. Regarding the type of the Kalman filter, since most biotechnical processes are non-linear, non-linear versions of the Kalman filter, specifically the EKF, are the most applied algorithm among other extensions of the Kalman filter. However the UKF is getting attention in recent years. The results in literature indicate that the UKF algorithms deliver more accurate estimates of the parameters and state variables compared to EKF algorithms.

In spite of the apparent success of Kalman filters for state and parameter estimation in lab-scale bioreactors, the integration of Kalman filters into industrial systems is not very widespread while most of the process models mentioned in literature consider noise-free ideal fermentations, whereas production-scale operations are corrupted by concentration gradients and disturbance. Accordingly, more efforts are required towards performing simulation studies in order to model and validate proper mathematical models associated with complex non-ideal bioprocesses.

Despite the numerous examples on state estimation methods for biotechnological processes in literature, the research on implementing Kalman filters for state

estimation in downstream processing remain rather limited. The advancement in state and parameter estimation methods in downstream processes leads to better knowledge of the location and concentration of the product and key contaminants, which are essential for process optimization and control.

So far most of the Kalman filter algorithms are implemented for monitoring fed-batch cultivations; however more attention is required for real-time implementation of the Kalman filter algorithms for controlling the feed rate and substrate production in these cultivations. Further efforts are also required towards implementation of state estimation methods in batch and continuous cultivations.

From the presented literature, it could be concluded that the non-linear extensions of the Kalman filter are powerful tools for state estimation in bioprocesses; therefore they could be used for digitalization of bioprocesses. Accordingly, in a case study, a digital twin of the baker's yeast batch fermentation process was developed by using a dynamic non-linear model of the process as well as an EKF algorithm. The proposed method gives the possibility to predict glucose, ethanol and biomass concentrations simultaneously from the only available infrequent online measurements of ethanol concentration. The accuracy of the estimated biomass and substrate production are in line with other studies which have also implemented an EKF algorithm for monitoring the baker's yeast cultivation [32, 49]. However, in our application the maximal specific growth rates on glucose and ethanol are also estimated. As a consequence, the rapid and precise estimation of these variables could increase the overall knowledge integration in the digital twin of the process.

Overall, the unique advantage of online monitoring and in general digital twins of bioprocesses is that they could play critical roles in bioprocess development such as supporting problem solving in manufacturing, reducing effort in setting up a control strategy and accelerating process performance by taking corrective actions automatically and in real time.

Appendix

Extended Kalman filter Matlab code: Online state prediction of batch yeast cultivations based on ethanol gas sensors.

```
%Initialization
clear; close all; clc;
sympref('AbbreviateOutput', false);
%Variable and parameter definition
%Symbols for symbolic math calculations
syms G E X P t                                  real
syms Y_gx Y_ge Y_ex mu1 mu2 K_M_G K_M_E         real
%Variables / Parameters
initX  = [2.5; 6; 0.2; 0.15; 0.08];             % initial state (Biomass,
% Glucose, Ethanol)
```

```
initP  = diag([0.1,0.02, 0.02,0.2,0.02]); % initial process estimation
% covariance matrix
init   = [initX; initP(:)];                % combined initial value
%vector
% for the odesolver
H     = [0 0 1 0 0]                        % observation matrix
H = 1×5
   0   0   1   0   0
Q = diag([0.001,0.001,0.001,0.001,0.001]) % process noise covariance
%matrix

R      = 0.05                              % measurement noise
%covariance matrix
K1     = 0.1;                              % Monod konstant glucose
K2     = 0.1;                              % Monod konstant ethanol

%estimated parameter values
Ygx    =0.15;                              % Yield glucose -> biomass
Yge    =0.34;                              % Yield glucose -> ethanol
Yex    =0.43;                              % Yield ethanol -> biomass
%Process model
%Monod terms
mue1 = mu1*G / (G+K_M_G);
mue2 = mu2*E / (E+K_M_E) * (1 - mue1/mu1);
%Model OD
dS = sym(X * [...
      ( mue1      + mue2)    ;...          % Biomass
      -mue1/Y_gx           ;...           % Glucose
      ( mue1/Y_gx*Y_ge - mue2/Y_ex) ;...  % Ethanol
      0;                                  % mue1
      0;                                  % mue2
     ]);
%Jacobian of Model with respect to state variables
F = jacobian(dS,[X,G,E,mu1,mu2])
P matrix
P = sym('P',[5,5])
dP = F * P + P*F'+Q
%Simulation / State prediction and filtering
%Replace all symbolic parameters with their respective numeric values
F  = subs(F, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
dS = subs(dS, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
dP = subs(dP, [Y_gx Y_ge Y_ex K_M_G K_M_E], [Ygx Yge Yex K1 K2]);
%Assemble all differential equations into a vector of 12 elements
%(3x state, 9x P)
OdeSys = matlabFunction([dS(:);dP(:)],'Vars',{t,[X; G; E; mu1; mu2;
P(:)]});
%load measurement values from file:

load BC2_eth_pred.mat                      % Ethanol sensor
%measurements
load BC2.mat                               % Offline values for

%Simulate the process from one ethanol gas measurement time to the next:
t0 = 0;
```

```
MC = zeros(0,5);          %store filtered states in these variables
SimState = zeros(0,5);
SimTime = [];
for i = 1:numel(timeE)
  tspan = [t0 timeE(i)];
  [T,state] = ode45(OdeSys, tspan, init);      % simulate / solve model

  PS      = state(end,1:5)';                   % predicted state

  MS      = ME(i);                             % measured state
  P       = reshape(state(end,6:end),5,5);     % process covariance
                                               % matrix
  K       = P*H'/(H*P*H'+R);                   % kalman gain matrix
  FS      = PS + K * (MS-PS(3));               % filtered state
  Pfilt    = P-K*H*P ;                         % filtered process
                                               % covariance matrix
  init     = [FS; Pfilt(:)];                   % new initial condition
  t0      = timeE(i);                          % new starting time for
                                               % next iteration
% Save intermediate states for plotting
  MC       = [MC;FS'];
  state(end,1:3) = NaN;
  SimState    = [SimState; state(:,1:5)];
  SimTime     = [SimTime; T];
end
%Results
%Plot the results in a presentable figure and save file to disk
f = figure("Position",[0,0,1600,640]);
subplot(1,2,1);
h = plot([0,time'],[initX(1:3)';M],'.','MarkerSize',20); % Plot
%measurements
set(h, {'color'},{'r'; 'g';'b'}); hold on;
h = plot(SimTime,SimState(:,1:3));            % Plot simulated values
set(h, {'color'}, {'r'; 'g';'b'});
plot(timeE,ME,'+b','MarkerSize',8); hold off; % Plot ethanol gas sensor
%values
ax = gca;
ax.FontSize = 14;
ax.FontName = 'Times';
ax.Position = [.05 .1 .4 .85];
ax.ActivePositionProperty = 'outerposition';
ax.GridLineStyle =':';
ax.GridAlpha = .7;
xlabel('time $/h$','interpreter','Latex',"FontSize",16);
ylabel('concentration $/\frac{g}{L}
$','interpreter','Latex',"FontSize",16); ylim([0 8]);
grid on; box off; grid(gca,'minor');
legend('Biomass offline','Glucose offline','Ethanol offline','Biomass
Kalman','Glucose Kalman','Ethanol Kalman','Ethanol gas
sensor','interpreter','Latex',"FontSize",12,"Color",[.9 .9 1 .9]);
subplot(1,2,2);
h = plot(SimTime,SimState(:,4:5));        % Plot mu values over time
ax = gca;
```

```
ax.FontSize = 14;
ax.FontName = 'Times';
ax.Position = [.55 .1 .4 .85];
ax.ActivePositionProperty = 'outerposition';
ax.GridLineStyle =':';
ax.GridAlpha = .7;
ytickformat('%.2f')
set(h, {'color'}, {'r'; 'k'});
xlabel('time $/h$','interpreter','Latex',"FontSize",16);
ylabel('$\mu$ value $/\frac{1}{h}
$','interpreter','Latex',"FontSize",16);
grid on; box off; grid(gca,'minor');
legend('$\mu_1$','$\mu_2
$','interpreter','Latex',"FontSize",12,"Color",[.9 .9 1]);
annotation("arrow",[.55 .97],[.1 .1])
annotation("arrow",[.05 .47],[.1 .1])
annotation("arrow",[.05 .05],[.1 .98])
annotation("arrow",[.55 .55],[.1 .98])
saveas(f,'KalmanPred.svg','svg');        % save copy of figure to file

%Calculate Errors
SimTime = SimTime + ((1:numel(SimTime))*1e-10)';
SimValues = interp1(SimTime,SimState(:,1:3),time);
SSE = sum((SimValues - M).^2);
RMSE = sqrt(SSE/numel(time));
SQT = sum((M-mean(M)).^2);
RSq = 1-SSE./SQT;
T1 = table('Size',[3,3],'VariableTypes',
{'double','double','double'},'VariableNames',
{'Biomass','Glucose','Ethanol'},'RowNames',{'RMSEP','Error
%','R²'});
T1(1,:) = num2cell(RMSE);
T1(2,:) = num2cell(SSE./(max(M)-min(M))*100);
T1(3,:) = num2cell(RSq)
```

References

1. Zobel-Roos S, Schmidt A, Mestmäcker F, Mouellef M, Huter M, Uhlenbrock L, Kornecki M, Lohmann L, Ditz R, Strube J (2019) Accelerating biologics manufacturing by modeling or: is approval under the QbD and PAT approaches demanded by authorities acceptable without a digital-twin? PRO 7(2):94
2. Kuchemüller KB, Pörtner R, Möller J (2020) Digital twins and their role in model-assisted design of experiments. In: Advances in biochemical engineering/biotechnology. Springer, Berlin. https://doi.org/10.1007/10_2020_136
3. Nargund S, Guenther K, Mauch K (2019) The move toward biopharma 4.0: In silico biotechnology develops “smart” processes that benefit biomanufacturing through digital twins. Genet Eng Biotechnol News 39(6):53–55

4. Luttmann R, Bracewell DG, Cornelissen G, Gernaey KV, Glassey J, Hass VC, Kaiser C, Preusse C, Striedner G, Mandenius C-F (2012) Soft sensors in bioprocessing: a status report and recommendations. Biotechnol J 7:1040
5. Schügerl K, Bellgardt KH (2012) Bioreaction engineering: modeling and control. Springer, Berlin
6. Schügerl K (2001) Progress in monitoring, modeling and control of bioprocesses during the last 20 years. J Biotechnol 85(2):149–173
7. Narayanan H, Luna MF, von Stosch M, Bournazou MNC, Polotti G, Morbidelli M, Butté A, Sokolov M (2019) Bioprocessing in the digital age - the role of process models. Biotechnol J 761. https://doi.org/10.1002/biot.201900172
8. Kalman RE (1960) A new approach to linear filtering and prediction problems. Trans ASME J Basic Eng 82:S.35–S.45
9. Wan EA, van der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: Proceedings of the IEEE 2000 adaptive systems for signal processing, communications, and control symposium (Cat. No. 00EX373). IEEE, pp 153–158
10. Matthews M (1990) A state-space approach to adaptive nonlinear filtering using recurrent neural networks. In: Proceedings IASTED Internat. Symp. artificial intelligence application and neural networks
11. Boulet G, Kerr Y, Chehbouni A, Kalma JD (2002) Deriving catchment-scale water and energy balance parameters using data assimilation based on extended Kalman filtering. Hydrol Sci J 47 (3):449–467
12. Krämer S, Grum M, Verworn HR, Redder A (2005) Runoff modelling using radar data and flow measurements in a stochastic state space approach. Water Sci Technol 52(5):1–8
13. Williams M, Schwarz PA, Law BE, Irvine J, Kurpius MR (2005) An improved analysis of forest carbon dynamics using data assimilation. Glob Chang Biol 11(1):89–105
14. Annan JD, Hargreaves JC, Edwards NR, Marsh R (2005) Parameter estimation in an intermediate complexity earth system model using an ensemble Kalman filter. Ocean Model 8 (1–2):135–154
15. Julier SJ, Uhlmann JK (1997) New extension of the Kalman filter to nonlinear systems. In: Signal processing, sensor fusion, and target recognition VI, vol 3068. International Society for Optics and Photonics, pp 182–193
16. Evensen G (1994) Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics. J Geophys 99(C5):10.143–10.162
17. Houtekamer PL, Mitchell HL (1998) Data assimilation using an ensemble Kalman filter technique. Mon Weather Rev 126(3):796–811
18. van der Merwe R (2004) Sigma-point Kalman filters for probabilistic inference in dynamic state-space models. Doctoral dissertation, OGI School of Science and Engineering at OHSU
19. Lee SC, Hwang YB, Chang HN, Chang YK (1991) Adaptive control of dissolved oxygen concentration in a bioreactor. Biotechnol Bioeng 37(7):597–607
20. Ghoul M, Dardenne M, Fonteix C, Marc A (1991) Extended Kalman filtering technique for the on-line control of OKT3 hybridoma cultures. Biotechnol Tech 5(5):367–370
21. Dubach AC, Märkl H (1992) Application of an extended kalman filter method for monitoring high density cultivation of Escherichia coli. J Ferment Bioeng 73(5):396–402
22. Gudi R, Shah S (1993) The role of adaptive multirate Kalman filter as a software sensor and its application to a bioreactor. IFAC Proc 26(2):249–254
23. Gudi R, Gray I, Shah S (1993) Multi-rate estimation and monitoring of process variables in a bioreactor. In: Proceedings of IEEE international conference on control and applications, IEEE
24. Albiol J, Robusté J, Casas C, Poch M (1993) Biomass estimation in plant cell cultures using an extended Kalman filter. Biotechnol Prog 9(2):174–178
25. Gudi RD, Shah SL, Gray MR (1995) Adaptive multirate state and parameter estimation strategies with application to a bioreactor. AICHE J 41(11):2451–2464
26. Petrova M, Georgieva O, Patarinska T (1995) State and time delay estimation of continuous microorganisms cultivation. Bioprocess Eng 12(1–2):103–107

27. Zorzetto L, Wilson J (1996) Monitoring bioprocesses using hybrid models and an extended Kalman filter. Comput Chem Eng 20:S689–S694
28. Holwill IJ, Chard SJ, Flanagan MT, Hoare M (1997) A Kalman filter algorithm and monitoring apparatus for at-line control of fractional protein precipitation. Biotechnol Bioeng 53(1):58–70
29. Hrnčiřík P, Náhlík J, Havlena V (1998) State estimation of Baker's yeast fed-batch cultivation by extended Kalman filter using alternative models. IFAC Proc 31(11):601–606
30. Ganovski L, Bliznakova M, Patarinska T (1999) State estimation of a Uricase production process with Candida utilis. Bioprocess Eng 21(3):273–277
31. Bogaerts P (1999) A hybrid asymptotic-Kalman observer for bioprocesses. Bioprocess Eng 20 (3):249–255
32. Hitzmann B, Broxtermann O, Cha Y-L, Sobieh O, Stärk E, Scheper T (2000) The control of glucose concentration during yeast fed-batch cultivation using a fast measurement complemented by an extended Kalman filter. Bioprocess Eng 23(4):337–341
33. Arndt M, Hitzmann B (2001) Feed forward/feedback control of glucose concentration during cultivation of Escherichia coli. IFAC Proc 34(5):403–407
34. Longhi L, Marcon S, Trierweiler J, Secchi A (2002) State estimation of an experimental bioreactor using the extended Kalman filtering technology. IFAC Proc 35(1):379–382
35. Patnaik P (2003) On the performances of noise filters in the restoration of oscillatory behavior in continuous yeast cultures. Biotechnol Lett 25(9):681–685
36. Cha Y-L, Hitzmann B (2004) Ultrasonic measurements and its evaluation for the monitoring of Saccharomyces cerevisiae cultivation. Bioautomation 1:16–29
37. Arndt M, Hitzmann B (2004) Kalman filter based glucose control at small set points during fed-batch cultivation of Saccharomyces cerevisiae. Biotechnol Prog 20(1):377–383
38. Arndt M, Kleist S, Miksch G, Friehs K, Flaschel E, Trierweiler J, Hitzmann B (2005) A feedforward–feedback substrate controller based on a Kalman filter for a fed-batch cultivation of Escherichia coli producing phytase. Comput Chem Eng 29(5):1113–1120
39. Patnaik PR (2005) The extended Kalman filter as a noise modulator for continuous yeast cultures under monotonic, oscillating and chaotic conditions. Chem Eng J 108(1–2):91–99
40. Rocha I, Veloso AC, Ferreira E (2006) Design of estimators for specific growth rate control in a fed-batch E. coli fermentation
41. Henry O, Kamen A, Perrier M (2007) Monitoring the physiological state of mammalian cell perfusion processes by on-line estimation of intracellular fluxes. J Process Control 17 (3):241–251
42. Soons Z, Shi J, Van der Pol L, Van Straten G, Van Boxtel A (2007) Biomass growth and kLa estimation using online and offline measurements. IFAC Proc 40(4):85–90
43. Klockow C, Hüll D, Hitzmann B (2008) Model based substrate set point control of yeast cultivation processes based on FIA measurements. Anal Chim Acta 623(1):30–37
44. Veloso AC, Rocha I, Ferreira E (2009) Monitoring of fed-batch E. coli fermentations with software sensors. Bioprocess Biosyst Eng 32(3):381–388
45. Jianlin W, Liqiang Z, Tao Y (2010) On-line estimation in fed-batch fermentation process using state space model and unscented Kalman filter. Chin J Chem Eng 18(2):258–264
46. Jianlin W, Xuying F, Liqiang Z, Tao Y (2010) Unscented transformation based robust kalman filter and its applications in fermentation process. Chin J Chem Eng 18(3):412–418
47. Bavdekar VA, Prakash J, Patwardhan SC, Shah SL (2011) Moving window ensemble Kalman filter for delayed and multi-rate measurements. IFAC Proc 44(1):11997–12002
48. Dewasme L, Goffaux G, Hantson A-L, Wouwer AV (2013) Experimental validation of an Extended Kalman Filter estimating acetate concentration in E coli cultures. J Process Control 232(2013):148–157
49. Popova S, Ignatova M, Lyubenova V (2013) State and parameters estimation by extended Kalman filter for studying inhomogeneous dynamics in industrial bioreactors
50. Sbarciog M, Coutinho D, Wouwer AV (2014) A simple output-feedback strategy for the control of perfused mammalian cell cultures. Control Eng Pract 32:123–135

51. Fernandes S, Richelle A, Amribt Z, Dewasme L, Bogaerts P, Wouwer AV (2015) Extended and unscented Kalman filter design for hybridoma cell fed-batch and continuous cultures. IFAC-Papers 48(8):1108–1113
52. Dewasme L, Fernandes S, Amribt Z, Santos LO, Bogaerts P, Wouwer AV (2015) State estimation and predictive control of fed-batch cultures of hybridoma cells. J Process Control 30:50–57
53. Zhao L, Wang J, Yu T, Chen K, Liu T (2015) Nonlinear state estimation for fermentation process using cubature Kalman filter to incorporate delayed measurements. Chin J Chem Eng 23(11):1801–1810
54. Krämer D, King R (2016) On-line monitoring of substrates and biomass using near-infrared spectroscopy and model-based state estimation for enzyme production by S. cerevisiae. IFAC-Papers 49(7):609–614
55. Simutis R, Lübbert A (2017) Hybrid approach to state estimation for bioprocess control. Bioengineering 4(1):21
56. Krishna VV, Pappa N, Rani SJV (2018) Implementation of embedded soft sensor for bioreactor on Zynq processing system. In: 2018 international conference on recent trends in electrical, control and communication (RTECC), IEEE
57. Krämer D, King R (2019) A hybrid approach for bioprocess state estimation using NIR spectroscopy and a sigma-point Kalman filter. J Process Control 82:91–104
58. Ritschel TK, Boiroux D, Nielsen MK, Huusom JK, Jørgensen SB, Jørgensen JB (2019) The extended Kalman filter for nonlinear state estimation in a U-loop bioreactor. In: 2019 IEEE conference on control technology and applications (CCTA), IEEE
59. Feidl F, Garbellini S, Luna MF, Vogg S, Souquet J, Broly H, Butté A (2019) Combining mechanistic modeling and Raman spectroscopy for monitoring antibody chromatographic purification. PRO 7(10):683
60. Lisci S, Grosso M, Tronci S (2020) A geometric observer-assisted approach to tailor state estimation in a bioreactor for ethanol production. PRO 8(4):480
61. Sonnleitner B (2013) Automated measurement and monitoring of bioprocesses: key elements of the M 3 C strategy. In: Measurement, monitoring, modelling and control of bioprocesses. Springer, Berlin, pp 1–33
62. Biechele P, Busse C, Solle D, Scheper T, Reardon K (2015) Sensor systems for bioprocess monitoring. Eng Life Sci 15(5):469–488
63. Vojinović V, Cabral JMS, Fonseca LP (2006) Real-time bioprocess monitoring: part I: in situ sensors. Sensors Actuators B Chem 114(2):1083–1091
64. Chhatre S (2012) Modelling approaches for bio-manufacturing operations. In: Measurement, monitoring, modelling and control of bioprocesses. Springer, Berlin, pp 85–107
65. Monod J (1949) The growth of bacterial cultures. Annu Rev Microbiol 3(1):371–394
66. Henson MA, Seborg DE (1992) Nonlinear control strategies for continuous fermenters. Chem Eng Sci 47(4):821–835
67. Jones KD, Kompala DS (1999) Cybernetic model of the growth dynamics of Saccharomyces cerevisiae in batch and continuous cultures. J Biotechnol 71(1–3):105–131
69. Contois DE (1959) Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures. Microbiology 21(1):40–50
69. Yousefi-Darani A, Paquet-Durand O, Babor M, Hitzmann B (2020) Model-based calibration of a gas sensor array for on-line monitoring of ethanol concentration in *Saccharomyces cerevisiae batch cultivation*. Biosyst Eng 198(2020):198–209
70. Galvanauskas V, Simutis R, Levišauskas D, Urniežius R (2019) Practical solutions for specific growth rate control systems in industrial bioreactors. PRO 7(10):693
71. Oisiovici RM, Cruz SL (2000) State estimation of batch distillation columns using an extended Kalman filter. Chem Eng Sci 55(20):4667–4680
72. Hashemi R, Engell S (2016) Effect of sampling rate on the divergence of the extended Kalman filter for a continuous polymerization reactor in comparison with particle filtering. IFAC-Papers 49(7):365–370

Adv Biochem Eng Biotechnol (2021) 177: 127–166
https://doi.org/10.1007/10_2020_153

Published online: 31 December 2020

The Challenge of Implementing Digital Twins in Operating Value Chains

Roman Werner, Ronny Takacs, Dominik Geier, Thomas Becker, Norbert Weißenberg, Hendrik Haße, Rudolf Sollacher, Michael Thalhofer, Bernhard Schumm, and Ines Steinke

Contents

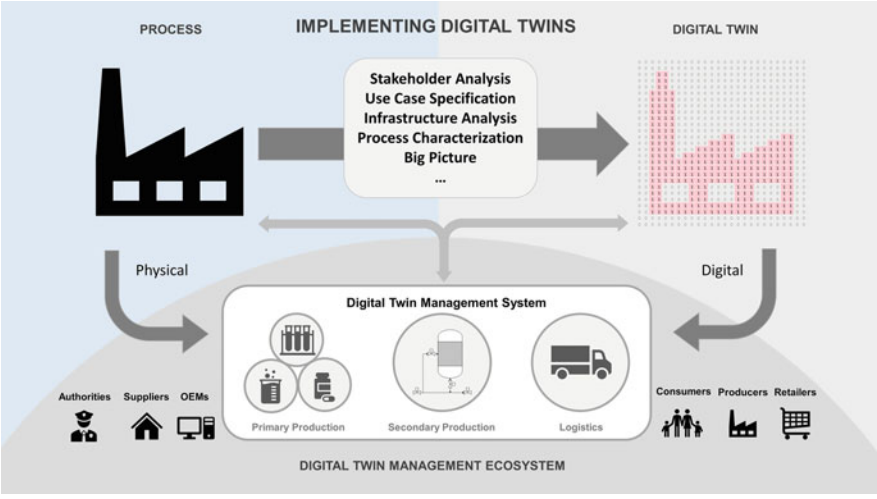
R. Werner, R. Takacs, D. Geier (✉), and T. Becker
Chair of Brewing and Beverage Technology, Technical University of Munich, Munich, Germany
e-mail: dominik.geier@tum.de

N. Weißenberg and H. Haße
Fraunhofer Institute for Software and Systems Engineering, ISST, Dortmund, Germany

R. Sollacher, M. Thalhofer, B. Schumm, and I. Steinke
Siemens AG, Munich, Germany

Abstract The concept of digital twins has become increasingly popular in recent years. To exploit their full potential, integration of systems and data across entire value chains is required. Implementing digital twins to newly built plants or production lines is challenging and even more complicated for currently operating production processes or factories. This chapter reviews and discusses strategies and tools to successfully implement digital twins into operating value chains in bioprocess and related industries. Furthermore, the implementation is exemplified with three recent case studies.

Graphical Abstract



Keywords Digital twin management systems, Ecosystems, Operating value chains, Standardization

1 Introduction

Industry 4.0 (I4.0), the new industrial revolution, refers to the emerging trends of digitization concepts for industries such as bioprocesses, food production, and operation value chains. Initiated by the German government, I4.0 comprises concepts and respective initiatives for digitalization and conception of modern goods manufacturing [1]. I4.0 includes various approaches to the digitization of new as well as the existing production processes. One central concept is the Digital Twin,

which represents the virtual copy of the existing physical objects [2], these physical entities are also called assets. Academicians have abbreviated Digital Twin as DT [2, 3]. DTs can represent almost any aspect of an associated asset, ranging from highly aggregated information to a detailed description of its components and performance; even simulation models of assets can be part of a DT. The final result depends largely on the use cases, which are addressed with the DTs.

Applying DTs to newly built plants or production lines is challenging and even more complicated in operating production processes or factories. Heterogeneous processes have various requirements, which result in different obstacles in establishing a working Digital Twin Management System (DTMS), used to integrate different DTs. Plants and their processes often evolve, which leads to complexity. Inhomogeneous and even incompatible systems result in connectivity problems and imprecise interrelations. Following Rosen et al. [4], DTs are not just a collection of virtual objects, but require interrelations, connections, and structure within a DTMS to leverage the full potential.

Connectivity, modularity, and autonomy are key enablers of DTs [4]. They improve process development, production planning, process intelligence, production execution, and the individualization of products and equipment. DTs connect the virtual networks and systems with the real world. Furthermore, full trackability and traceability, which are essential for food safety, become feasible. Artificial intelligence (AI)-enabled concepts and technologies lead to state-of-the-art production concepts and the establishment of production intelligence. Finally, DTs, together with a DTMS, support companies to prepare for the challenges of I4.0. Kritzinger et al. [3] categorize a DT into a Digital Model and a Digital Shadow based on the level of integration. DTs comprise an automatic data flow between physical and virtual objects. Their functionality depends on the accuracy of the underlying semantic description, the assignment of relevant information, and a well-designed structure in the DTMS. Therefore, suitable techniques and standards must be applied. For example, the *Asset Administration Shell* (AAS) is a domain-independent standard of the German *Platform Industrie 4.0*, which specifies how to construct DTs, namely their data models and their interfaces, that allow efficient interaction in Industry 4.0 scenarios. It is being developed as a standardized software interface of any physical assets.

This chapter explains the strategies and tools to implement DTs into operating value chains or industrial processes and demonstrates its successful implementation with three case studies by analyzing the value chain, corresponding stakeholders, and the process as well as the primary infrastructure. From the initial step with an analysis of the stakeholder and DT goals to the final implementation, the chapter includes all necessary steps for successful execution. Figure 1 provides a graphical overview of the general implementation pathway which is outlined in the different sections of this chapter.

Notably, the status quo associated with the development of the future desired digitalized structure (DTs and DTMS) is illustrated. A physical model (physical picture) and a data model (virtual picture) of the considered value chain are also included showing the interrelations and interfaces among stakeholders. The physical

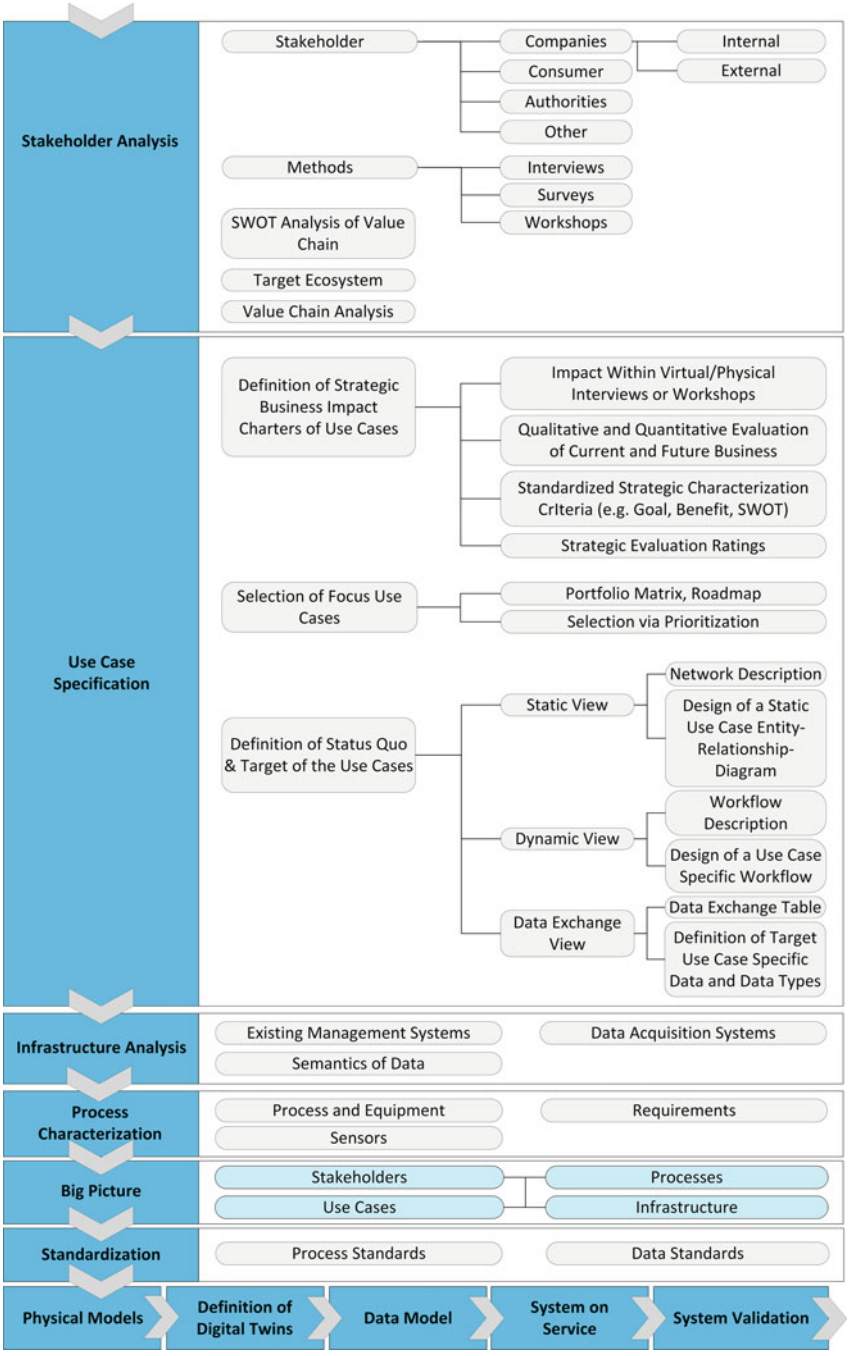


Fig. 1 Step-by-step process of implementing digital twins into an operating production or value chain

model includes the original equipment such as machines, sensors, and actuators as physical assets. The data model represents the advancement and comprises all data points that can be acquired from the relevant components of the physical production process. It also includes all management systems in the operating production process that are part of the DTMS. Notably, the data model also consists of the individual order of the DTs. In addition, this chapter reviews relevant standards that support the establishment of an operating DTMS and highlights future opportunities.

2 Industrial Bioprocesses and Corresponding Value Chains

Biological transformation processes determine our daily life. Nature developed various principles that are also applicable to industrial production processes. Process engineers use different bioprocess-related, procedural techniques, e.g., fermentation, to transform substrates to the desired consumer good. According to Liu [5], bioprocesses are present in the following industrial areas: biomaterials, health, biology, process industry, biofuel, and food. In particular, the production of food, pharmaceuticals, and biotechnological goods requires elaborate bioprocesses.

For these reasons, bioprocesses are elementary parts of value chains. A value chain adapts the production steps of a specific product to its ecosystem or the company's complete manufacturing process if only the corresponding parts of the value chain are considered. In order to depict the industrial perspective in a reasonable way, it is essential to evaluate the value chain as a whole. Thus, not only the bioprocesses should be considered but all other unit operations involved. For the production of bioproducts (e.g., food), several production steps are required. Exemplary processes in the food industry are the production of beer or yogurt, where fermentations are fundamental for their production. This chapter outlines in Sect. 7 several case studies that belong to the food industry and corresponding logistics.

A complete overview of the production environment is necessary to guarantee full interconnections of DTs within certain company areas or even among individual companies. In particular, tracking and tracing must consider all elements of the value chain.

3 Analysis of Operating Value Chains

A DTMS implementation usually starts with the value chain analysis of the associated ecosystem (Fig. 1). A value chain is analyzed by identifying the stakeholders, the operating production lines, what physical and virtual components need to comprise the DTMS, and the types of requirements. Subsequently, the use cases and targets of the DTMS, including the interfaces and connections between the stakeholders, need to be specified. Furthermore, the stakeholders' technical and virtual infrastructure must be analyzed. This helps in defining the process borders,

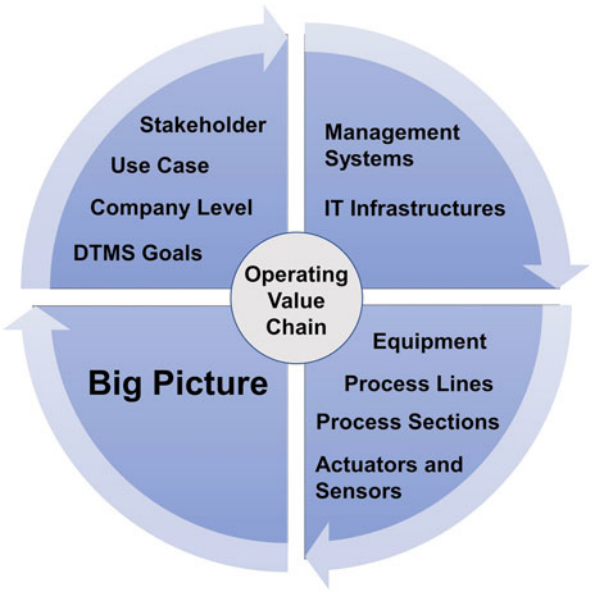


Fig. 2 Overview of value chain analysis

identifying the desired data points in the production process, and leading to overall standardization. In-depth knowledge of these factors is necessary to create a big picture of the entire structure of the DTMS (Fig. 2).

3.1 Stakeholders Analysis

Different stakeholders participate in the value chain and exert their influence. These members or groups perform numerous tasks, are responsible for particular tasks in the DTMS, and benefit from the system and associated tools. The stakeholder analysis depends on environmental conditions such as regional requirements and infrastructural conditions of a country (e.g., developing countries). The literature [6–10] outlines several considerable requirements and provides approaches for conducting a stakeholder analysis in compliance with the preexisting conditions.

Regarding a DTMS, stakeholders are distinguishable in company's internal as well as external groups. Internally, stakeholders include employees and managers with a given responsibility (e.g., production manager, warehouse manager, and quality manager). Externally, stakeholder groups consist of other companies (e.g., raw material suppliers), governmental organizations (e.g., food authorities), and customers. Therefore, the following specific aspects must be considered for DTMS:

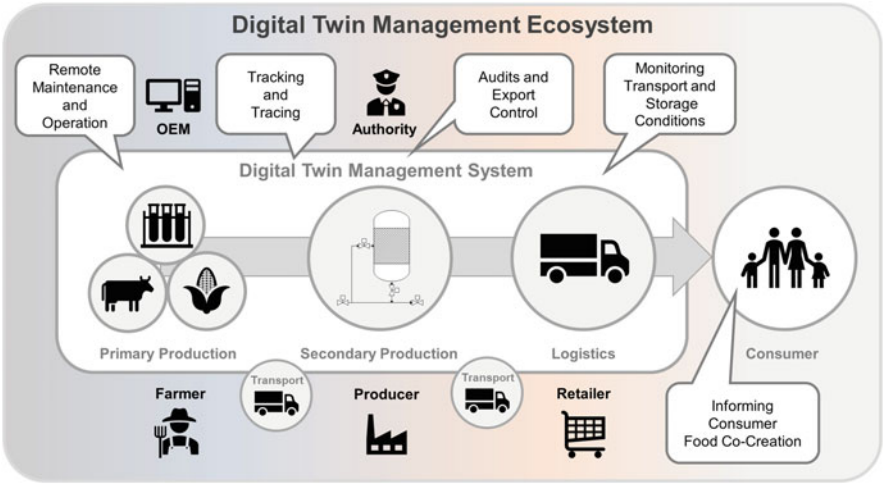


Fig. 3 Sharing digital twin data with stakeholders in a Digital Twin Management Ecosystem and corresponding tasks

- Scope of the DTMS: Is it designed for internal purposes only, or will it also involve the production processes of other companies in a business-to-business solution (B2B)?
- What companies, associations, and authorities will take part in it?
- Who is involved in the production process in the company internally (e.g., quality manager, operators, and logistics manager)?
- Is a business-to-customer (B2C) solution considered?

Following stakeholders' identification and distribution of relevant tasks, a stakeholder network is conceptualized. Figure 3 shows an exemplary bioprocess value chain, relevant stakeholders, and associated tasks.

3.2 Use Case Specification

Use cases refer to specific application examples of the planned system (e.g., the DTMS) and pursue specific goals. Additionally, they are ideal for testing the functionality of the system and pave the way for further applications. However, the actual reliability must be compared with all influencing parameters and always be subjected to a critical review [11, 12]. Use cases comprise defined targets and represent a system's behavior according to the users' requirements for reaching those targets. These users are individuals or groups and represent stakeholders. Each use case has distinct tasks and aims that must be achieved by applying a specific procedure. Use cases indicate DTs' responsibilities and are always named

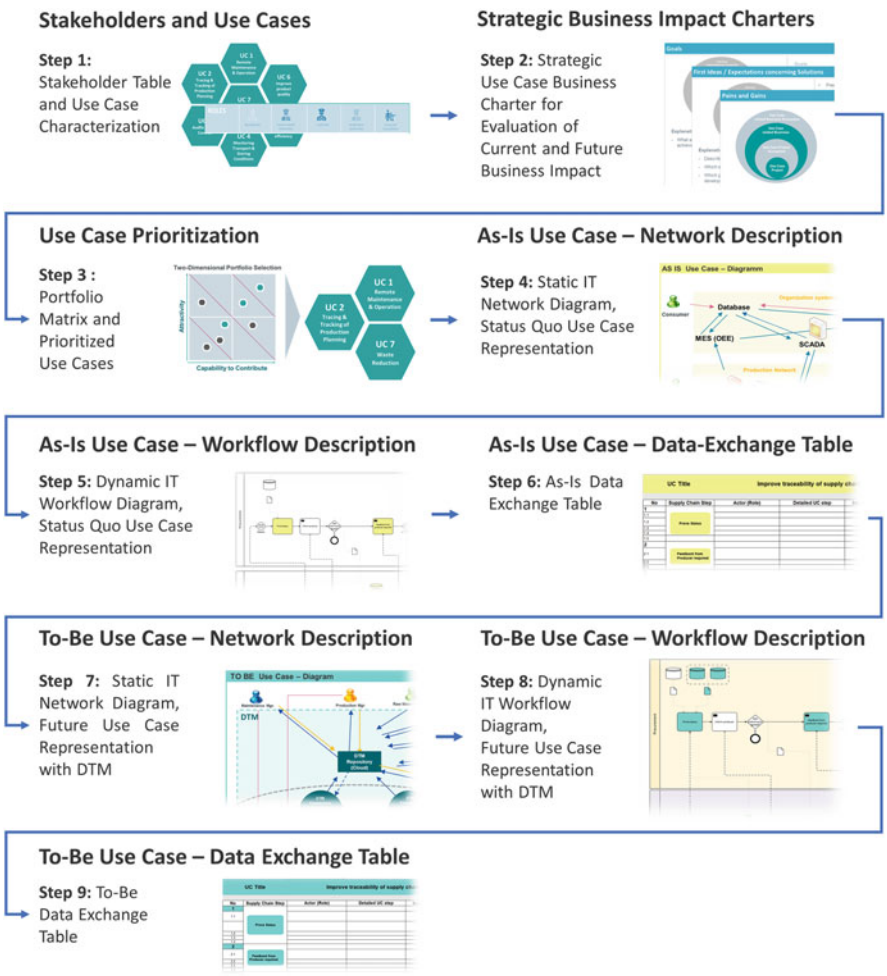


Fig. 4 Process for identification, selection, and description of use cases

according to their primary goals. Concerning whole value chains, the product trackability would be a model use case for implementing DTs. Besides, use cases illustrate all likely scenarios when a stakeholder acts using the respective system.

Use cases for DT-based application scenarios are described for the current situation ("As-Is") or for the future ("To-Be") and are designed with static (e.g., network diagrams) or dynamic (e.g., workflows) representations. They provide the basis for deriving requirements and the subsequent solution design (software development).

For defining use cases at a strategic and operational level, the following steps and Fig. 4 describe a detailed systematic approach. This systematic approach provides a

comprehensive, condensed, multi-view, and consensual basis for quality-assured requirements engineering as a basis for all subsequent tasks, from data modeling to the development of DT-based software functions.

Step 1: Identification of Stakeholders and Use Cases

Based on a general understanding of the use cases, company internal stakeholders as well as external stakeholders of the ecosystem are identified. They contribute to the methods described in the steps 2–9.
Methods: Virtual or face-2-face interviews, surveys, workshops with design thinking, and creativity sessions
Results: Stakeholder table and use case characterization (short description)

Step 2: Definition of Strategic Business Impact for Use Cases

The current and future business impact for the identified use cases is analyzed.
Methods: Qualitative description and quantitative evaluation of strategic business impact is analyzed within interviews and workshops. Hereby standardized strategic issues (e.g. goals, benefits, and SWOT) and measurable criteria for strategic ratings are used.
Results: Strategic use case business charter

Step 3: Prioritization and Selection of Focus Use Cases

The use case candidates from step 1 and 2 are prioritized based on the evaluation results from step 2. Most attractive use cases are selected as focus use cases by all stakeholders.
Methods: Prioritization and selection methods, e.g. scoring methods, AHP (analytic hierarchy process)
Results: Portfolio matrix and prioritized use cases

Step 4: Definition of Status Quo (As-Is) Use Cases – Static view of IT assets

Definition of a static IT network diagram of As-Is focus use cases showing all relevant IT components (e.g., hardware, software), roles, actors, and their data-input/output relations.
Methods: Design of a static use case entity-relationship-diagram
Results: Static IT network diagram: As-Is Use Case Representation

Step 5: Definition of Status Quo (As-Is) Use Cases – Dynamic view of IT assets

Development of a workflow diagram of As-Is focus use cases showing logical sequence of activities of roles or actors, and their data input/output relations
Methods: Design of a use case specific workflow
Results: Dynamic IT workflow diagram: As-Is use case description

Step 6: Definition of Status Quo (As-Is) Use Cases – Data exchange view

Specification of data (e.g., data types, data attributes) exchanged by relevant actors.
Methods: Interviews and workshops on further data concretization

Results: As-Is data exchange table with detailed data specifications

Step 7: Definition of Future (To-Be) Use Cases – Static view of IT assets

Definition of a static IT network diagram of To-Be focus use cases (utilizing Digital Twin) showing all relevant IT components (hardware, software), roles, actors, and their data-input/output relations.
Methods: Design of a static use case entity-relationship-diagram
Results: Static IT network diagram: To-Be use case description

Step 8: Definition of Future (To-Be) Use Cases – Dynamic view of IT assets

Development of a workflow diagram of To-Be focus use cases (utilizing Digital Twin) showing logical sequence of activities of roles or actors and their data input/output relations
Methods: Design of a use case specific workflow
Results: Dynamic IT workflow diagram: To-Be use case description

Step 9: Definition of Future (To-Be) Use Cases – Data exchange view

Specification of data (e.g. data types, data attributes) exchanged by relevant actors.
Methods: Interviews and workshops on further data concretization
Results: To-Be data exchange table with detailed data specifications

In summary, the collaboratively elaborated results of all steps of this systematic approach provide a comprehensive and condensed basis (Big Picture, see Sect. 3.5) for a systematic methodology from requirements engineering and data modeling to the development of Digital Twin based software functions. Investment in infrastructure for digitalization must consider several highly ranked use cases: The return on investment may be high, and the investment for implementing an additional use case is usually not as significant as for establishing the infrastructure for the first use case. Several attractive use cases can be achieved only if DT data is shared among different stakeholders (Fig. 3).

3.3 Infrastructure Analysis

DTs are a sophisticated link between the virtual and the physical world, but their implementation is application-specific [3]. However, the following are the minimum requirements for the implementation:

- The digital environment (e.g., server and network structure)
- A standardized process
- Extent of automation level following the minimum requirements
- Definition of a minimum data structure for communication
- Expert, plant, machine, and product knowledge

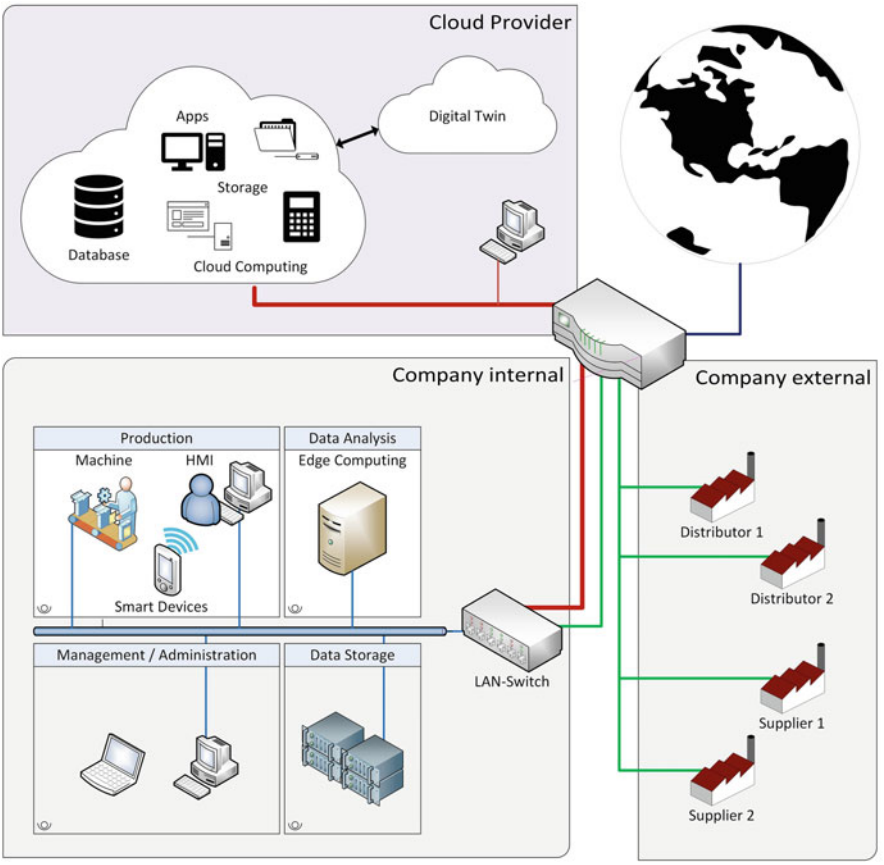


Fig. 5 A simple concept of essentials for IoT, edge computing, cloud computing, and Digital Twins

- Enabler technologies for physical system integration
- Definition of sensor, operational, and transactional data
- Definition of system granularity
- Connectivity of manual and automated tasks

Data sampling and communication must consider these minimum requirements in the planning phase (Fig. 5). The production line is linked to other departments such as administration, shipping, and other on-site facilities such as servers and databases, e.g. by using cloud services. The communication might address external company locations, sales partners, distributors, or suppliers. The all-inclusive DT is located in the cloud and uses its services (e.g., for data storage, databases, and service apps). The cloud and the DT (asset core) are linked to all aforementioned components. The cloud provider is responsible for the final hosting in the cloud.

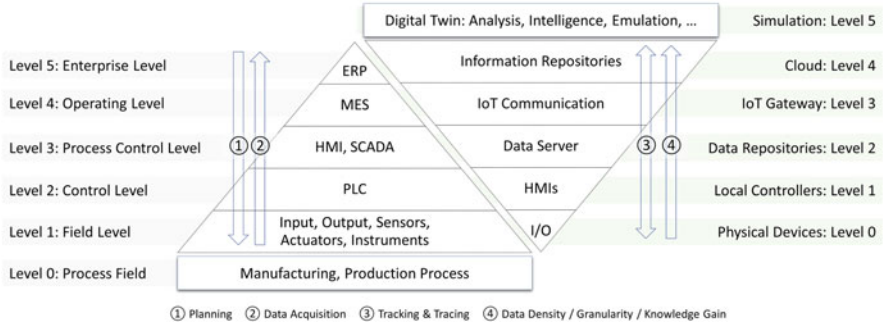


Fig. 6 Architecture of automation pyramid (Siepmann model based on DIN EN 62264 [15, 16]) in context to Digital Twins (adapted from [17])

Furthermore, the communication structure is critical for extensive networking within the company, enabling real-time data traffic between the DT and its counterparts (physical assets). A DTMS requires bidirectional communication. This interrelation can be implemented through a wired or wireless network and radio standards such as 5G. Notably, the 5G standard is widening possibilities in creating local area data networks to use mobile and rapid information exchange inside plants using smart mobile devices [13]. A well-established network is essential for the DT to share real-time data with physical assets such as control modules, sensors, machines, and human–machine interfaces (HMI).

Data communication must follow common language, universal communication, and standardized structure. The existing industrial communication protocols commonly used for process monitoring and control can be used for data communication. In practice, owing to the numerous companies acting worldwide, more than one format or communication protocol is commonly used in one production plant. Here, converters may translate the data into a dominating protocol.

HMIs are the link between the human operator and the DT. They integrate human intelligence and skills to improve efficiency [14]. By defining particular transfer points in the physical systems, HMIs can be implemented into the DT, which adds less automated or even offline modules to the system. Every manual input affects the system's consistency and should, therefore, be avoided.

The granularity of the DT can be determined within the network structure and connectivity of the physical assets. A simple DT is limited to a particular physical asset such as a product, administration task, single machine, or equipment. Increasing the granularity of the DT increases the number of linked physical assets. Therefore, DTs can be used to observe a single production process, or, according to the automation pyramid of Siepmann [15], DTs can be linked to field, control, process control, operating, and enterprise level. With increasing granularity, the amount of data, network traffic, and computational calculations increases. These details result in higher investment costs and thus lead to an improved process resolution. A holistic DT is the objective of modern intelligent systems.

Figure 6 shows the holistic approach of a cyber-physical system connected to the DT according to the automation pyramid of Siepmann [15, 16]. On the left, 0–5 describe the automation pyramid separated into strictly hierarchical company levels: 0 is the sensor/actuator level, characterized by simple and rapid data sampling; 1 is the field level, the interface to the production process; 2 is the control level; 3 is the process control level; 4 and 5 are the plant management and company levels, respectively, where production planning, production data acquisition, and order processing are conducted. The pyramid of the DT is illustrated on the right. Levels 0–1 contain the physical assets, where level 0 includes actuators, sensors, and equipment and level 1 provides additional functionality such as HMIs for the DT. Level 2 contains communication and data servers. Level 3 is the gateway between the physical assets and the DT, and all the other structures connected to the system. Levels 2 and 4 communicate over the gateway, whereby level 4 supports cloud-based databases with information from the physical assets and DT. Levels 0–4 allocate the required infrastructure for the DT. Level 5 consists of the digital core asset, the DT with all necessary apps such as simulation, emulation, and modeling. With increasing levels in the automation pyramid, the data flow increases. This correlation offers extensive networking, thus enabling the full capability of the DT.

The connection of levels 1–3 in the DT pyramid is achieved in production lines through hierarchy, whereby three levels of networking are distinguished: unit, system, and system of system (SoS) level [18]. The unit level is the smallest element and can comprise machine equipment, materials, or sensors. The system level is a combination of multiple unit levels, which can communicate and control each other by field bus, Ethernet, or 5G. SoS is composed of multiple system levels as in the collaboration of multiple production lines and combines all data of the production lifecycle.

3.4 Process Characterization

Process characterization requires a detailed analysis of all physical production equipment (including sensors) and the procedures to identify all the production parts that are linked with the product, or relevant for the supply of utilities (e.g., electricity, steam, and gases), or affect the process quality indirectly (e.g., cooling systems and logistics). The production departments and floors relevant for implementation into the DTMS must be defined. Within these individual sections, identification of equipment and defining of the process borders ensure the assignment of the process to the most appropriate subsections. Moreover, the defined process borders help to model the correct process sequence in the DTMS and reduce its overall complexity (Fig. 7).

In process engineering, a process consists of a sequence of several chemical, physical, or biological unit operations [19, 20]. Each unit operation aimed to convert, transport, or store raw materials or intermediates. Combining all process steps ensures the production of the desired good and its packaging. Industrial production

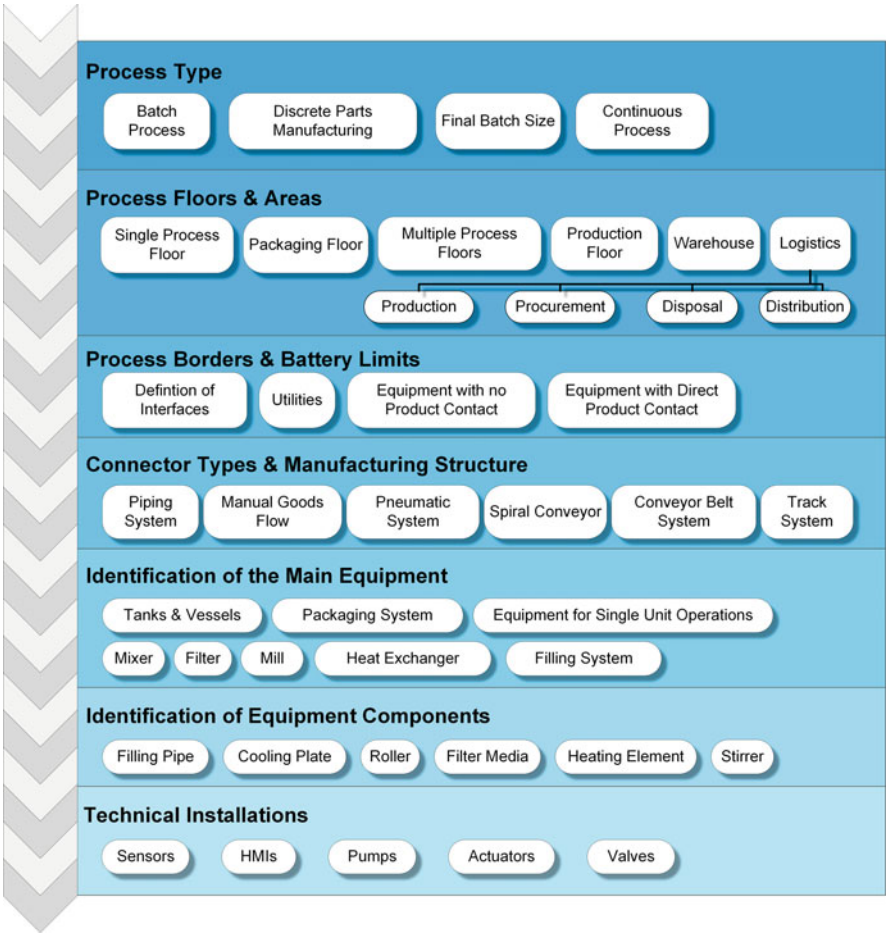


Fig. 7 The different steps of process analysis with exemplary components

processes can be categorized as batch, continuous, and discrete manufacturing [21]. A batch process represents a discontinuous production process that consists of various unit operations. In contrast, a continuous process runs through production without interruption, and raw materials and intermediates are processed by passing each production stage. Discrete parts manufacturing illustrates the production processes that create a specific quantity of intermediates. Once the specific quantity is achieved, further processing occurs. Accurate classification of the process is essential for selecting the appropriate standards and identifying the key process steps (unit operations).

Following the identification of all the relevant process components and their classification into "with product contact," "utilities-relevant," and other markings, their semantics must be defined. This terminology highlights the need for correct

naming or numbering of the components, description of the process task, classification of the process relevance, and other designations that help in understanding their role in the DTMS. For these instructions, all the existing process descriptions are beneficial. Exemplary records are flow charts, SCADA-charts (Supervisory Control and Data Acquisition), illustrations from management systems, and results of laboratory management systems.

3.5 Composition of the Big Picture

The big picture provides the DTs' structure and architecture and comprises the DTMS as well as its connectors, functions, and the necessary physical model and data model. It is an integrated complete high-level vision of the "To-Be" system, depicting the desired future operation. It considers current trends (e.g., digitization), historical social perspective (e.g., consumer demands), economic perspective (e.g., influence of the market), and forecast (e.g., influence of artificial intelligence; effects of robots) as well as own objectives.

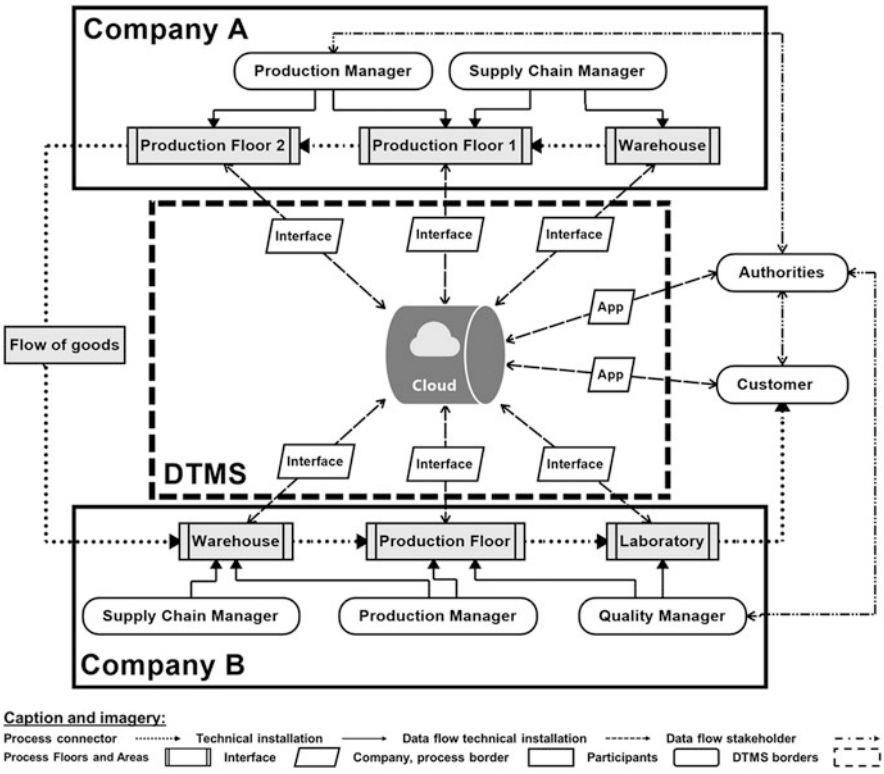


Fig. 8 Big picture of an exemplary DTMS with stakeholders

Here, the DTMS forms the core element consisting of the cloud storage, the relevant stakeholder interfaces, and the apps. In particular, the apps provide special tools (e.g., analysis and predictive tools) that enable the full potential of the system. Furthermore, it is important to link all participating elements to define the interrelations within the process borders and between the individual parties. Its creation also supports the comprehension of interrelations between stakeholders and inherent assets. Figure 8 illustrates an exemplary big picture consisting of two different companies by showing connections and data flows.

4 Standardization and Generation of Digital Twins

A DTMS merges multiple production systems and data points, whose origin plays an essential role because these data sets are of different quality and comprehensiveness. The combination of various production lines, independent plants, and even different companies results in different kinds of data sets. In the case of extensive data sets, a reasonable junction in the DTMS is unfeasible. This aspect is confirmed by Weyer et al. [22], who researched on modular, multi-vendor systems in the context of Industry 4.0. Therefore, there is a high demand for standardization and relevant norms. Furthermore, the use of a DTMS must comply with the corresponding legislation, for example, hygienic requirements or work safety concepts. These aspects are essential as a DTMS may also function as the basis for other concepts.

Standardization is a critical dimension of sustainable and efficient engineering. Standards specify the design features of equipment, create comparability between manufacturers, and help customers receive a safe product that complies with existing legislation. According to the *German Institute for Standardization* (DIN, Deutsches Institut für Normung e.V.) [23], standards are the universal language of engineering and facilitate free movement of goods. In a DTMS, standards facilitate a comparison of different data sets, production lines, plants, or even companies. They help create individual DTs and their application in a DTMS. Finally, standardization also provides the language that facilitates the communication between DTs. Lu et al. [24] emphasized the importance of standards for DTs by analyzing the state-of-the-art manufacturing domains.

4.1 Overview of Existing Standards

International, continental, and national organizations define the standards for the public. Globally, the *International Organization for Standardization* (ISO) for technical standards and the *International Electrotechnical Commission* (IEC) for standards with electrotechnical relevance function as the roof organizations. Currently, over 160 national standardization committees are members of the ISO [25]. The complexity of a DTMS, especially in B2B or B2C relations, challenges

standardization as new technologies are emerging. However, national and international standardization organizations release new drafts for standards continuously. The ISO has been developing a series of standards for DTs. The series will include four different parts, definitions, suitable reference architectures, digital representation of physical manufacturing elements, and information exchange [26–29], and define the framework of DTs. Another ISO standard under development focuses on the visualization elements of DTs [30]. These standards are expected to gain significance for implementing DTs in the future.

Standardization of DTs requires distinction between process and data standards. Process standards must include definitions of physical objects such as equipment, machines, or connectors (e.g., pipes). Production types must be defined, and company-related properties standardized. Here, the *International Society of Automation* (ISA) and the *American National Standards Institute* (ANSI) offer suitable standards for production processes. Notably, the three standard series ISA-88, ISA-95, and ISA-106 include appropriate methods and tools for DT standardization [21, 31–46]. The standardization process also supports the analysis of the basic properties of a production line and its transferability into a DTMS.

With regard to data sets, different levels of the data hierarchy must be defined. Following the *German Electrical and Electronic Manufacturers' Association* (ZVEI, Zentralverband Elektrotechnik- und Elektronikindustrie e. V.) [47], the description of data sets emerges from the following aspects:

- Data type (e.g., real, bool, and array) and data format (e.g., XML and JSON)
- Data source (e.g., alarm value and measured value)
- Data semantics (e.g., manual or automatic process)
- Data display (e.g., numerical and curves)
- Aggregated illustration of data (e.g., faceplates and complex diagrams)
- Functional integration of data (e.g., controller and HMI link)

This ontology-driven approach helps the operator to understand the type of data visible, its origin, and meaning. The DTMS requires a semantic web structure that helps distinguish between DTs. Management execution systems (MES) can be combined with DTs, which lead to a full MES-driven approach. The requirements of management and control of the quality (ISO 9001), the energy (ISO 50001), and occupational health and safety (ISO 45001) can extend DTMS appropriately and create a comprehensive production system [48–50].

General management systems, as well as food safety, are feasible in a DTMS. Here, requirements, limit values, or legislation are applicable to DTs. Food safety management systems (e.g., ISO 22000), similar concepts and measures, such as HACCP (hazard analysis and critical control points), are also associated with the DTMS [51]. These supplements help monitor the hygiene and food safety status in the production, train employees with customized instructions, and avoid or at least react promptly in case of a food safety event. Moreover, authorities and audit organizations are likely to participate in DTMS. The ISO facilitates corresponding bodies and audits in this field [52]. If a DTMS focuses on tracing and tracking, ISO suggests principles and basic requirements [53].

4.2 Process Standardization

Given that a DTMS can be local or global, it is essential to identify and list all its relevant components. Typically, this process analysis includes a complex mix of various companies, plants, process types, equipment designs, connectors, and sensors. This section outlines how this heterogeneity can be standardized. First, this study helps to achieve a general application of the various processes into a DTMS. Second, an appropriate standard of communication must be established in all associated processes. The ISO offers suitable standards that are regularly updated. It is necessary to identify what processes will be implemented into the DTMS (see Sect. 3.4).

The main focus is on the process type. The ANSI/ISA-88 (or S88, SP88) covers batch processes and contains four interdependent parts. The IEC and national standardization organizations have published corresponding versions (e.g., IEC 61512-1 and DIN 61512) [54, 55]. The ISA-88 shows the possibilities of structuring procedural systems and equipment in levels. A level organization helps in greater flexibility and increased performance of the considered batch process. Part 1 [21] includes the terminology as well as the definitions of the entire production process and its modeling possibilities. Part 2 includes the structures of data and key communication [31] (see also Sect. 3.3). Part 3 [32] treats product recipe models and their method of definition. Part 4 [33] comprises the recording of batch production. Other standards belonging to the ISA-88 explain the implementation (e.g., packaging equipment), as well as recipe formats, and are also useful for developing DTMS [41, 56].

For a DTMS's basic construction and architecture, part 1 is the most critical standard for batch processes. The physical model can be derived from this standard. This model represents the physical structure and the essential connections of the DTMS. Hierarchical levels create a structure that forms the entire production process. From top to bottom, the physical model grows holistically – groups of production parts in a lower hierarchy form a part of the subsequent higher level. Thus, the entire production process is illustrated, and a logic structure of relevant processes with the corresponding subgroups emerges. Any DT can be derived from the emerging modeled structures; that is, each part of a hierarchy can also represent its own DT.

Furthermore, virtual relations, which lead to the data model, are based on the physical model. The physical model illustrates the rigid interrelations of batch production. Figure 9 gives an overview of the selected terminology and levels of the ISA-88, the assignment to specific physical production components, and the adaption in a DTMS.

Following Fig. 9, the structure and process-related architecture of the DTMS can be derived. The linkages between distinct DTs must be defined. For example, the link between the two DTs is achieved through pipes or conveyor belts. If there are valves that can monitor and record the time (e.g., timestamp of valve switching), the batches are trackable and traceable.

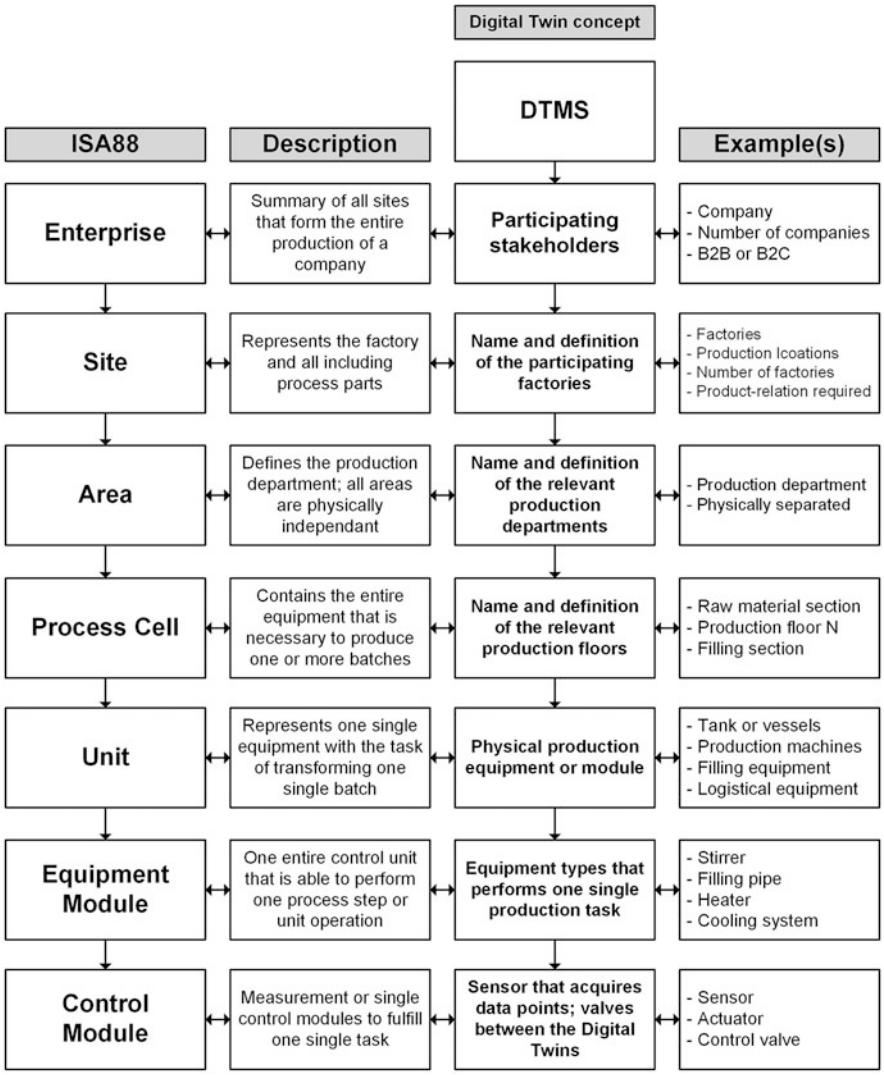


Fig. 9 Application of the ISA-88 physical model to standardize a process and implementation into a DTMS

The ISA-88 focuses on process cell, unit, equipment module, and control module. Terminology and background of enterprise, site, and area are defined in ISA-95. This standard deals with the integration of the corporate and operational management levels and is based directly on ISA-88. Besides batch production, it is also applicable to continuous and discrete production processes. The ISA-95 consists of seven parts including models, definitions, object model attributes, activity models, possibilities of management operations, business-to-manufacturing transactions, messaging

services, and an alias service model [34–37, 39, 40, 45]. The international version of this standard is the IEC 62264, which is based on the works of ISA. With the help of ISA-95, various management systems can be applied to the DTMS. It also helps plan the access of various stakeholders or operators within the production process. In combination with ANSI/ISA-101.01-2015 [46], the possible methods of implementing HMIs help in the efficient control of the DTMS on-the-machine or with mobile devices. An additional standard combines the implementation of ISA-88 and ISA-95 [44].

For continuous production, the ISA-106 offers an in-depth alternative [42, 43]. In comparison to ISA-88 and ISA-95, this standard series offers elaborate physical models that help organize even complex processes. For example, it defines a variable state model as illustrated by a user with high flexibility.

4.3 Data Standardization

Industry 4.0 factories have physical machines, or components called assets, connected to software that visualize the entire production line or make own decisions. An asset is an organization's physical entity having either a perceived or an actual value [57]. For bioprocesses, these are machines or different plant facilities and their parts.

Interoperability is the basis for I4.0 and ensures open and plural markets. It is characterized by open standards. The German *Platform Industrie 4.0* (PI4.0) develops pre-competitive concepts and solutions for I4.0, implements them, and participates in international standardization processes through more than 10 international cooperation [58]. The PI4.0 AAS specification [59] is the basis for interoperability and a DT standard. The intention is to become the central, standardized "integration plug" of any asset to digital ecosystems, composed of multiple DTs. Using an AAS, all relevant assets speak a common language, which eases integration. Any physical item that provides relevant data may become an asset and in turn get an AAS. Therefore, it is considered as a data standard as well. For example, the operating data of a plant and the production process can be standardized throughout their life cycle by building AAS DTs, which may be integrated into a DTMS.

The data exchanged or made available via an AAS is described in a modular, manufacturer-neutral format with formally described semantics. It does not prescribe what data is provided, but how they are provided. The generic AAS data model is defined using Unified Modeling Language (UML) class diagrams. Using an AAS, the DTs data is organized into submodels. Therefore, the main classifications are as follows:

- *Submodels*, which are either predefined entirely or may be described using a standard pattern. An AAS may have any number of submodels.
- *Properties*, which can be used to define the *submodels.*

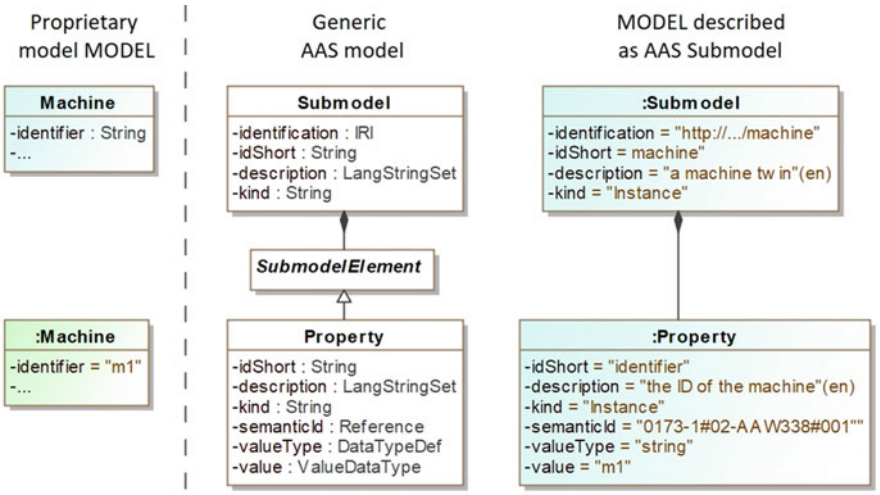


Fig. 10 Simple AAS submodel construction sample

Figure 10 illustrates the principal structure of submodels in the AAS standard and the method of developing an AAS submodel for any model. The left panel displays a simplified proprietary, that is, a nonstandard data model (named *MODEL*) of a machine's DT. Its instantiation is shown below, describing a concrete machine *m1*. Only one attribute is displayed here to focus on the mechanisms of translating this model to an AAS submodel.

The middle panel of Fig. 10 shows the simplified *generic* AAS model. The right panel shows that this model is instantiated to build the given proprietary model. Each submodel has an Internationalized Resource Identifier (IRI) [60], an *idShort* (often being the last part of the IRI), *descriptions* (several pairs of language tags and strings in that language are allowed), and a kind (being *Type* or *Instance*). A submodel is composed of *SubmodelElements*. Different subclasses exist, such as *Property*, *File,* and *Collection* (the latter are not shown).

For all attributes of the given model, an AAS *Property* can be added to the AAS *Submodel*. A *Property* is a name–value pair with additional metadata. Here the semantic annotation using its attribute *semanticId* is crucial for automatic interpretation. In this example, an International Registration Data Identifier to an attribute defined by the eCl@ss [61] standard is used. The eCl@ss dictionary contains properties for product descriptions and service descriptions based on standardized data formats conform to IEC 61360 [62]. Alternatively, an IRI referencing a standard property of well-known ontologies is usable (e.g., https://schema.org/identifier). A third alternative is to use own IEC61360 conformant so-called *ConceptDescriptions*, stored within an AAS.

The last attributes are *valueType* and *value*. A value can only be set if *kind* = *Instance*. This context indicates that types and their instances appear

similarly, and only the attribute *kind* and the attribute *value* differ. *DataTypeDef* is any XSD datatype [63], and *ValueDataType* is a value of such type.

Assets may be composed of other assets. In this case, a composite AAS is constructed, which has a bill of material listing its parts. These assets may be co-managed within the comprising AAS, or they may be self-managed, that is, they all have their own AAS. Therefore, complex AAS structures can be built, which may reflect the physical asset structuring.

The data model and the corresponding data can be serialized (i.e. automatically transformed to), for example, as extensible markup language (XML), JavaScript object notation (JSON), resource description framework (RDF), or open platform communications-unified architecture (OPC-UA). The standard also defines all these mappings. An AAS can be entirely transported as a so-called AAS extension (AASX) package. Furthermore, a representational state transfer API (REST-API) of an AAS is currently being defined and specified by PI4.0. It will be defined as part 2 of the AAS specification [59], but is work in progress, and intends to standardize the AAS interfaces.

Another module in the AAS information model is the attribute-based access control (ABAC) based security model, which protects the AAS, e.g. its REST-API. For each subject (role or user), it can be specified which object (submodels or even properties) the user is allowed or denied to read or to modify using expressions over attributes of the subject, the object, and the context.

The AAS specification does not provide methods to describe how a physical asset can be connected with its AAS. Anything is allowed here, hidden to the AAS users. Therefore, the AAS homogenizes the assets' diversity in the real world by providing a standard ecosystem plug.

4.4 Data Sharing Standards

DTs are a mature instrument for data collection and integration that can bridge media disruptions between distributed systems [64]. A DT forms a central knowledge base within a company and contributes to enhance business processes [64]. Although the benefits of DTs are not limited to their use for internal company processes, they are mainly used to monitor the internal processes. In this context, DTs are a suitable instrument for sharing data with various stakeholders [65], which companies find it increasingly important as data represents a strategic resource with economic value [66].

In principle, the difference between exchange and sharing of data must be recognized. Exchange of data occurs only in terms of vertical cooperation between companies, where the optimization of value or supply chains is one of the objectives. This type of data exchange is common and based on standards developed since the introduction of electronic data exchange in the 1980s. Data sharing, however, describes the vertical and horizontal collaboration between companies to achieve common goals; for example, predictive maintenance through collaborative data

sharing, in which both the company providing the data and that using it mutually benefit through improved services and databases [59].

The shared DT is necessary for implementing collaborative data sharing. It is based on the fundamental concepts of a general DT, which are characterized by the integration of various data formats from distributed data storage and the description of data with meta-information [67]. Following the definition of data sharing, a shared DT describes the extension of the archetypical characteristic of a DT by the functions of interoperable and sovereign use in collaborative networks. This extension includes the standardized data model, which in turn includes the uniform description of interfaces. The respective data model must enable manufacturer-neutral and cross-company interoperability.

A shared DT is further characterized by the emphasis on security concepts and, in this context, especially the concepts for data policy enforcement. Access control is an essential tool for data policy enforcement for DTs and restricts the access to functionalities and the asset model information of the DT [68]. However, data sovereignty must be considered in the use of DTs in collaborative networks. It is the ability of a natural or legal person to exercise exclusive self-determination over the economic asset data [59] and based on the data policy enforcement on usage control. In contrast to traditional access control, usage control regulates future data usage by adding restrictions [69]. Access control, therefore, manages the rights to collect data, while usage control determines how the recipient can use the data [70, 71].

Within an enterprise, data security, accountability, and transparency are defined. If data leaves the enterprise, additional security is required. For shared DTs, it must be ensured that the company providing the data always retains sovereignty. Therefore, the *International Data Spaces Association* (IDSA) [72] has developed a reference architecture, which enables the secure and sovereign exchange of data between trustworthy parties. It defines a technical infrastructure and a semantic set of rules for data exchange and data usage in ecosystems. DIN SPEC 27070 [73], based on the Industrial Data Spaces (IDS) reference architecture model, is the first global and interoperable standard.

The AAS is a virtual digital and active representation of an asset and renders a standardized I4.0 component in an I4.0 (eco)system composed of such building blocks. For multilateral data exchange, the AAS offers an appropriate basis for interoperability between the actors mentioned earlier through its upcoming standardization. The *Fraunhofer Gesellschaft* is currently working on the combination of IDS and AAS (see Case Studies in Sect. 7). The AAS-REST-API will be realized as an IDS Data App, which can be offered in an IDS App Store if necessary. IDS messages contain AAS-compliant data with IDS references to resources available via IDS. The ABAC of the AAS is synchronized with IDS Contracts and AAS subjects with IDS Participants. The IDS is more focused on communication (data in movement) and provides general interfaces but does not specify payload formats, whereas the AAS standards are more concrete. In complex and automated I4.0 scenarios, new legal issues arise, currently being investigated in the legal testbed project [74] in cooperation with IDSA and PI4.0.

A combination of both concepts, the IDS and the AAS, likely ensures interoperability between various stakeholders and respects individual data protection needs. Both are industry-neutral approaches, also applicable to production systems. Their combination is a reasonable basis for a shared DT. In particular, interoperability in this context refers to the likelihood of simple migration to various cloud and edge providers. The GAIA-X project [75] aims to enable such cross-cloud usage. In general, GAIA-X forms a networked and provider-neutral data infrastructure that enables secure storage (data in rest), sovereign exchange, and collaborative use of data and services.

5 Integration of Models and Data Sources into a DT-Compatible Platform

Integrating different data sources is a major advantage of implementing DTs. For example, a product's DT typically requires information from not only enterprise resource planning (e.g., batch id and recipe), but also process control (e.g., the used amount of supply material and produced product and resource consumption for the product). Data can be integrated by storing information from different sources on a specific platform, such as a cloud as well as an on-premise platform. An example of a cloud-based platform is MindSphere™ [76], which has been used in the EIT Food project "Digital Twin Management" (Sect. 7.1) [77]. The main advantage of using a cloud-based platform is that there is no need for resources to maintain the corresponding IT infrastructure and data backups and that the data is available everywhere. Platforms need to include a user and access management as well as IT security measures for the exchange of information.

A use case specification for a digitalization project may reveal the lack of data and interfaces. This is particularly significant for brownfield installations with low-level automation or connectivity and requires investments in the automation network, additional sensors, and engineering of additional data points. Connecting various data sources with the platform can be supported with suitable devices and software modules. Automation components such as controllers need to connect to the platform, as well as to a REST-API. Software tools require a connection to other sources such as SAP systems.

Once the data is available on a platform, it can be used for visualization and analysis. These apps can be created and provided as a service by any provider. However, data access and analysis are simplified mainly by using a suitable data model, providing a meaningful semantic description of the data on the platform.

DTs can also be used to increase the transparency of value chains. This is achieved by sharing parts of DTs with business partners or authorities. Data ownership must be respected by such a solution, for example, preventing data users from accessing or even manipulating an owner's data without consent. An excellent

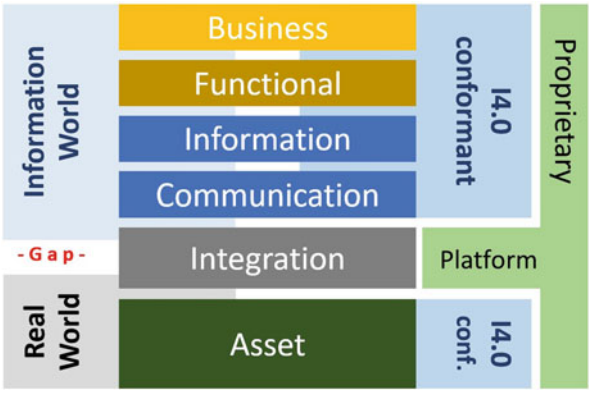


Fig. 11 RAMI 4.0 layers adopted from PI4.0 [79]

introduction to the different roles and some useful rules for data sharing can be found in the EU code of conduct on agriculture data sharing [78].

For successful integration in terms of I4.0, the Reference Architecture Model Industry 4.0 (RAMI 4.0) is a well-established guidance for implementation. It unifies the most relevant aspects of the I4.0 architecture and represents the holistic life cycle of assets. The RAMI 4.0 supports a uniform structure, wording, entire scope of I4.0, and complex relationships in small and clear parts. The RAMI 4.0 model consists of three dimensions: hierarchy, product life cycle, and architecture. Architecture is subdivided into six layers: business, functional, information, communication, integration, and asset. In the information layer, different data models can be integrated (e.g., submodels of asset administration shells), which get their data from the lower levels (i.e., communication or asset), while the functional and business layer provide the processing logic. Applied to bioprocesses, e.g. kinetic models can be integrated with other data sources, and the functional layer can combine them accordingly.

Figure 11 displays the layers of PI4.0 architecture reference model RAMI 4.0, which spans from real world to the information world. Between both worlds, there is a gap in the current PI4.0 standards: The asset layer is the lowest layer in RAMI 4.0 and represents the assets of the physical world, such as a machine in a production environment. The integration layer in RAMI 4.0 acts as a link between the real and the information world and is not yet standardized as the higher layers. This gap can be covered by the assistance of appropriate platforms.

6 Risk and Hurdles for a DTMS Implementation

Prior to the integration of a DTMS into operating plants, some major hurdles and risks must be addressed. Section 3 and 4 explain and discuss several plant adaptations scenarios in which risks could occur. According to Singh et al. [80] depending on the aspect to be considered, the challenges or risks can be classified into several groups: engineering, commercial, technology, and data challenges.

The engineering challenges and risks require standardization and hardware adaption (see Sect. 4). It is necessary that all physical assets communicate with the DTMS in one language. The integration of the DTMS with the automation systems of an existing plant may require additional often unforeseen efforts, in particular if these automation systems involve a proprietary solution or are lacking corresponding interfaces. Increasing the complexity of the systems related to the implementation of a DTMS is also a challenging task. To minimize risks, integration work must start very early, so that problems that could arise can be resolved in time. A detailed work plan must take foreseeable risks into account and provide a suitable buffer in terms of effort and time.

Commercial issues can include data sharing difficulties with the stakeholders that are part of the value chain. On the one hand, it offers a high level of synergy effects and numerous benefits, but on the other hand, information sharing and ownership could raise some challenging technological issues. In this context, Singh et al. [80] mainly addressed company policies, the way of thinking of involved stakeholders, and cultural differences. These difficulties and occurring issues can reduce the effectiveness and overall benefit. Customs control regulations could also cause delays and possibly additional costs in delivering the DTMS to the partners. Having to take into account multiple food safety regulations across the world may exceed not only the available resources for the project but also the complexity of the solution. This can be avoided by focusing on highly relevant market segments and/or regions.

Technology challenges may increase costs and implementation time, e.g. for CPS (cyber-physical system) implementations. An industrial CPS (ICPS) connects the cyber and virtual world. Notably, the life cycle of ICPS components needs to be clarified [81]. Often heterogeneous systems need to be integrated using different technologies; they implement different standards or even have proprietary interfaces. Moreover, communication or computation bottlenecks may occur when vast amounts of data are collected from the automation systems. The technical requirements of the new primary framework for the DTMS require detailed planning.

A more fundamental challenge in the DTMS implementation is the data itself. This requires intensive support through the highly complex and external communication across process and company borders. Secure system integration is a challenge for any company, especially when interfacing other companies' systems. Untrusted interfaced systems outside the company are a security risk. All information exchange should follow the common policies and possible threats have to be considered. The information flow should be controlled to limit attack vectors like identity thefts or privacy and security breaches. Information should also be validated and filtered. A standardization of the system integration process based on the definition of security requirements helps reduce these risks.

Internal data security must be redesigned based on these new demands. The existing systems must be adapted to upcoming or future requirements and to new security issues [82]. Data access control and even data usage control may be required when data sovereignty needs to be enforced. It needs to be supported appropriately by all involved systems. Moreover, the resulting new high-level of networking poses a particular risk. Data loss can range from image damage to loss of company trade

Table 1 Selected examples of risks, hurdles, and corresponding solutions

Step	Hurdles/current state	Consequence/solutions
Stakeholders analysis	Internal/external stakeholders; security risks by external stakeholders	Face-2-face interviews, surveys, workshops with design thinking, and creativity sessions
Use case specification	Analyze current and future business impact for use cases, prioritize them, define data and workflow including all relevant components, both for current and for desired situation	Interviews and workshops, qualitative description and evaluation of strategic business impact, SWOT, scoring methods, AHP, business process model and notation (BPMN), UML
Infrastructure analysis	Define minimum sensor, operational, and transactional data in sufficient granularity and define enabler technologies for physical system integration	Select integration technology, e.g. AAS, IDS, MindSphere
Process characterization	Analyze all physical production equipment and procedures, categorize them by Fig. 7, e.g. batch, continuous, and discrete.	How to support manual tasks by HMIs How to analyze the data (e.g., use of AI)
Composition of big picture	Integrated complete high-level vision of the "To-Be" system, depicting future operation	Abstraction, considers, e.g., current trends, social and economic perspective and own objectives
Process standardization	Many process standards, modeling detail	Select one, e.g. BPMN, structured according to ISA-88
Data standardization	Many data standards, modeling detail	Select one, e.g. UML, structured according to PI4.0 AAS, serialized as XML or JSON
Data sharing standardization	Data exchange or collaborative data sharing, access control and usage control	DIN SPEC 27070, ABAC, IDS Usage Control

secrets or intellectual property. This can be a challenging step, due to unclear internal and external communication channels at the beginning of the planning process. In the planning phase, all potential weak points need to be considered.

Further data challenges include concepts like data acquisition, type, transfer points, streams, database handling, big data storing, and data security. One potential risk is that the data collected from plants or products are not adequately representing the same concept. This can lead to, e.g., information available in digital twins being incomplete or not up-to-date, and not being sufficient for, e.g., getting reliable results from a root cause analysis. One possible solution is an overview of the treatment of food safety issues, which helps identify relevant information, and allows for early measures to provide this information, e.g., by installing new sensors and capturing new relevant data. By early and detailed planning, this risk can be reduced. Especially these points show the challenges of a DTMS implementation. Nevertheless, they are not unknown problems, and therefore they can be minimized by proper and detailed planning in advance (Table 1).

7 Case Studies

This section presents three case studies for implementing DTs in operating production lines or entire value chains. The work was carried out in different research projects in which the authors were actively involved.

7.1 Digital Twin Management: Implementation of a DTMS in an Operating Production Process

The first case study originates from the EIT Food project "Digital Twin Management" (DTM, 2018–2019) and was adapted from previous works [77, 83]. A DT structure and a DTMS were implemented into the value chain of pudding production, including the production of the required flavors. Thus, two different companies and corresponding plants required a suitable connection within the DTMS. The recipe for the pudding product included skimmed milk, flavors, sugar, fat, and other ingredients. A producer of flavors manufactured the flavor ingredients and a dairy plant was responsible for the actual pudding production. Figure 12 illustrates the corresponding big picture, which includes the physical model and the data model. Internally, in both companies, quality managers, warehouse managers, and production managers aimed to participate in the DTMS. In addition to the two companies, the stakeholder analysis identified three other major parties: authorities, customers, and logistics. Notably, the authorities and customers benefit from the DTMS via apps delivering suitable production data.

Process and structural characterization were first realized for the producer of flavors. Moreover, an MES and data storage system existed in the process line. The process type was batch related and comprised three different process sections: warehouse, production, and storage. Each production step represents an individual unit operation: juicing, filtering, and rectifying. According to the standards explained in Sect. 4, they are declarable as units. In addition, various sensors that measure temperature, pressure, and weight were identifiable. The system also registered timestamps of several process starts and ends and was confirmed via HMIs. These data points were declared as control modules according to the selected standards. Manual transport helped connect between the units, and timestamps registered the layover of a batch at a particular unit.

The dairy plant was also batch related and was equipped with various management systems (MES, LIMS), historical data storage, and partially automated production facilities. Five different process sections existed: warehouse, milk storage, pudding production, cream production, and packaging. Pudding production was targeted for the DT implementation. This process includes four steps: ingredient mixing, homogenization, pasteurization, and packaging. The cream production consists of fat storage, cream whipping, and cream storage. Both process sections include various physical assets such as sensors (e.g., temperature, pressure, and pH)

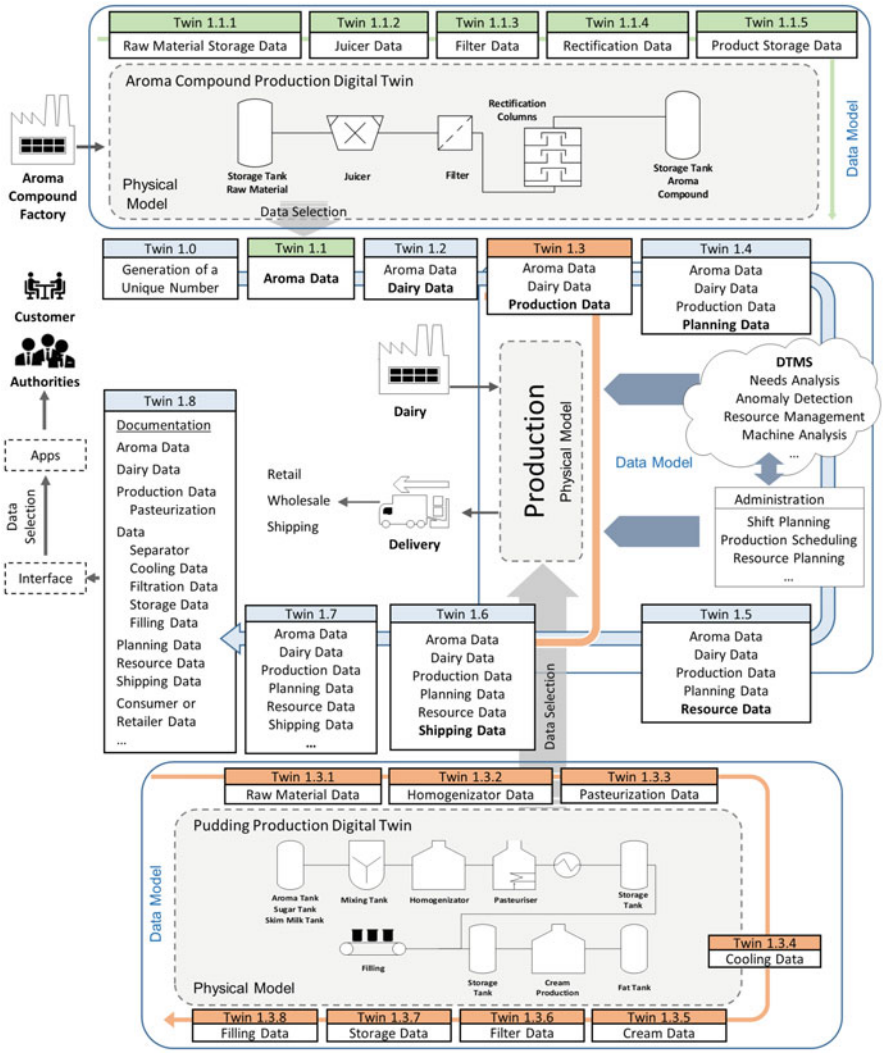


Fig. 12 Big picture of the DT implementation in the supply chain of dairy production (adopted from [77])

and actuators (e.g., motors and valves). The valves connected the individual sections defining timestamps for the product section transfer. Furthermore, some extensions could connect incompatible and non-standardized machines and sensors to the DTMS. HMIs were also considered with a comprehensive documentation option and included in the DTMS.

Two physical models were derived from the analysis of the companies and processes. Given that the project aimed at tracking and tracing, only the equipment

and components with direct product contact were considered. Following further analysis of the sensory and the structural conditions (especially IT-related structures), the two corresponding data models with all relevant data points for the DTMS were created. These virtual models also comprise the actual DTs and their data. In terms of the physical model, the DTs are classified in the correct production order and with a suitable semantic designation. Owing to the progress in the production within the value chain, the DT is continuously growing in terms of data equipment, as new, relevant data points are added in each further process step related to the physical model. Hence, the first DT in the value chain is the data poorest, while the last DT receives the most and all relevant data points. Subsequently, all the models were connected via appropriate interfaces and the respective analyzed stakeholders were involved. The overall structure of all the components finally results in the big picture, which shows the DTMS, the structure of the DTs, and the stakeholders involved (Fig. 12). The big picture results in individual structures for each value chain and must always be created anew. Furthermore, it provides a complete overview of the architecture of the DTMS and apps for the stakeholders. It is the crucial step in the implementation of a DTMS and has a considerable impact on its functionality.

In this case study, DTs aimed to ensure full traceability and trackability along the value chain. Besides, the connectivity between the companies, plants, production areas, and process cells was reached. These measures support the use of PI4.0 potentials and the stakeholder participation. The connection of both companies considered by the full big picture creates one DTMS that enables the complete reconstruction of the processes.

7.2 Organic Supply Chains: Implementation of a DTMS for Vegetable and Beef Supply Chains

The approach of the DTM project (Sect. 7.1) is also used and extended in the EIT Food project "Organic Supply Chains" (OSC, 2019, ongoing) [84], which builds a MindSphere™ DTMS for vegetable and beef supply chains. The main objective is to prove the organic status and safety of the produced food. The supply chains extend from ordering and planning over processing to delivering to retailers. They connect farmers, food producers, transporters, and retailers. Figure 13 presents an overview of such a supply chain.

In the DTM project, all involved companies were project partners. Thus, all data points and systems were accessible by them, and the complete production process could be covered. In contrast, in the OSC project, only the retailers are currently the project members, resulting in limited data and source system access. In the OSC project, there is no concrete physical model of the farms with detailed specifications of work centers and units for which DTs are requested. Instead, complex business process models and notation models are used to define all the process steps. The

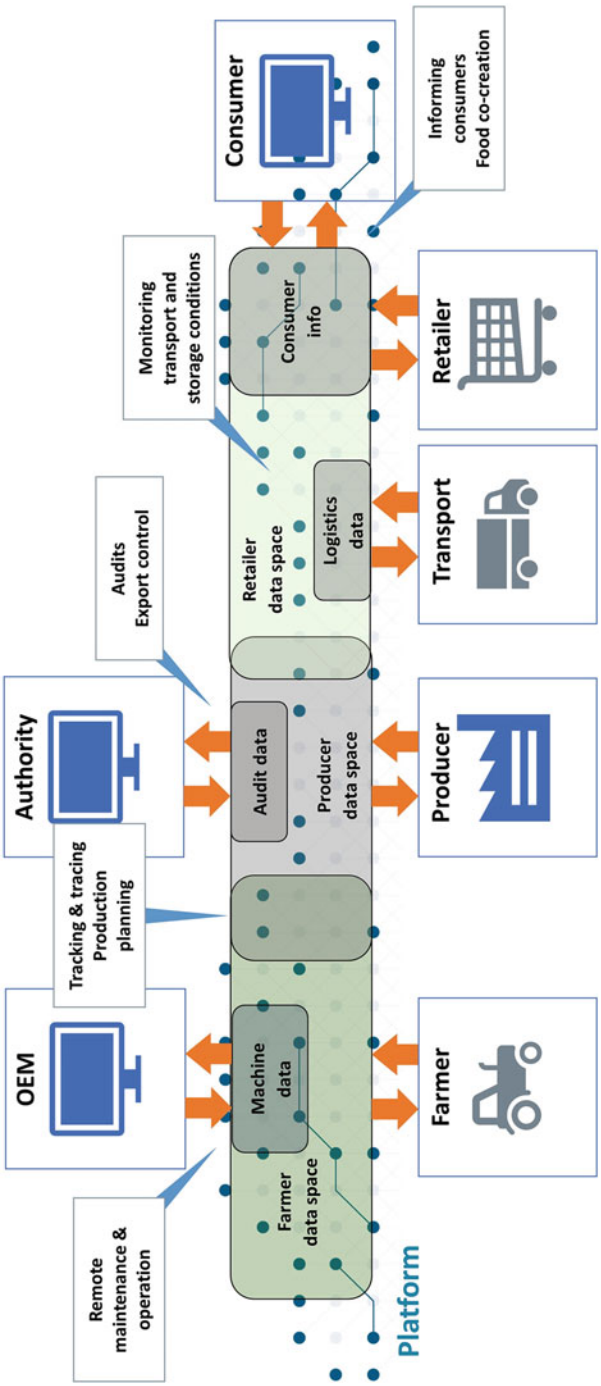


Fig. 13 Sharing digital twin data with stakeholders in an agri-food value chain enables many use cases

relevant activities/work units and even the work centers were identified following ISA-88. Organic supply chains have strong linkages and communication between the process participants and the majority of the steps are manual. Moreover, the data is complex, and the processes involve various communicating partners. Therefore, detailed UML models are used in the OSC project and structured according to ISA-88 hierarchies.

7.3 Shared Digital Twins

The Fraunhofer cluster of excellence "Cluster of Cognitive Internet Technologies (CCIT)" [85] studies cognitive technologies for the industrial internet. Researchers from multiple disciplines are developing key technologies for various levels along the value chain, from sensors and intelligent learning methods for data processing up to cloud technologies. Its *Forschungszentrum Data Spaces* (FDS) focuses on concepts and technologies for sovereign industrial data exchange based on the IDS. An integral aspect of the FDS is the cross-institutional cooperation between numerous Fraunhofer Institutes collaborating on projects. One of these projects relates to the integration of IDS Connectors and AAS and, accordingly, with the combination of the information models and the security concepts of both approaches. It aims to develop a shared DT based on the current and future standards for interoperability and data sovereignty.

RIOTANA® (Realtime IoT Analytics) is a domain-independent IoT architecture and DT that processes raw sensor data into key process indicators (KPIs) in real-time, developed by Fraunhofer ISST [86]. It consists of sensor modules attached to arbitrary assets (e.g., forklifts), which transmit their data via Message Queuing Telemetry Transport (MQTT) to a backend, which can combine the sensor values and calculate KPIs. The combination of a proprietary DT such as RIOTANA® with IDS and AAS leads to a shared DT complying with the corresponding standards. The data can be accessed by multiple participants in the ecosystem, while the data owner retains full control over the data. The owner can define who can access and use the data and for what purposes the data can be used. By using IDS, its usage control mechanisms can be applied.

Figure 14 shows the architecture of the combination: on the right panel, as sample assets, three forklifts *f1–f3* are shown with attached RIOTANA® sensor modules, which transmit their values to the RIOTANA® DT. The latter is now an IDS data app in a service container and used to implement the standard AAS-REST-API, which also is an IDS data app acting as the AAS wrapper for RIOTANA®. In this case study, there is a composite forklift fleet asset with co-managed forklift assets.

The combination of AAS and IDS results in an architecture that requires the mapping of the data models and security concepts: IDS messages contain AAS-compliant data with references to IDS resources. The AAS-ABAC security concept is combined with IDS contracts, which protect those resources. The submodels of the AAS are protected by both mechanisms and may be subject to

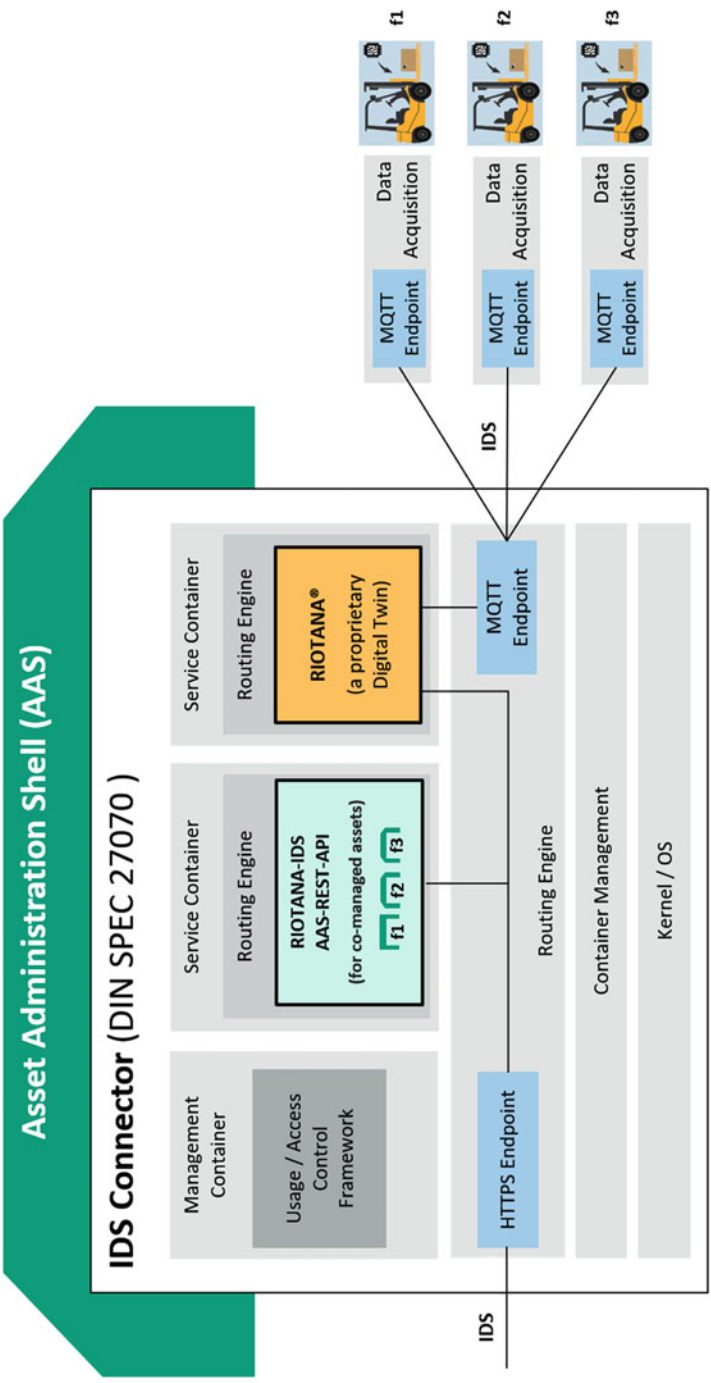


Fig. 14 Architecture of the RIOTANA-IDS-AAS

IDS usage control. This approach is independent of RIOTANA®; it can be used to convert any proprietary DT into a standard (AAS) sovereign (IDS) sharable DT by providing an IDS-AAS wrapper. This also holds for proprietary DTs of production processes and their parts and products. The use of established concepts and approaches for interoperability and data security is, therefore, the basis for the design of shared DTs.

8 Summary and Outlook

In I4.0, new types of cooperation have replaced rigid value chains. The boundaries between country, industry, and company are being blurred. Plants, machines, and products communicate autonomously in digital, globally connected networks. Therefore, dynamic and self-optimizing ecosystems will emerge, which comprise several decentralized stakeholders. In digital ecosystems, all relevant physical assets must be integrated into the I4.0 environments to interact with each other. Everything is mapped digitally and uniformly, ensuring transparency along the entire value chain. Moreover, there are no instances of central control, but several decentralized combined "building blocks." In an open ecosystem, no stakeholder may assume a monopolistic position. The core elements are DTs. There are various solutions: exemplary platforms provide extensive solutions for integrating asset data into DTs, ranging from hardware to software, but the solution is proprietary. On the other hand, standardization approaches, such as the AAS, a key concept of the German PI4.0, exist. It, however, does not provide or standardize any means to connect the asset to the AAS and hence any mechanism can be used.

DTs and DTMS generate voluminous amounts of valuable data accessible and processable, which enable several novel applications. The concept of DTs helps determine the current state of a product along the entire product life cycle. Therefore, the design of the DT must provide the communication between the DTs and with complex architecture (e.g., management, administration, and planning services). Individual machines or entire production lines are also represented by distinct DTs (e.g., in the form of an AAS). However, the abundance of data often requires a strict definition of system boundaries or requires data usage control in the global digital ecosystem.

In bioprocesses and related industries, DTs enable a full digitalization with all associated concepts and tools. Not only complete monitoring of the whole value chain becomes possible but also tracking and tracing of the individual steps/units. Bioprocess unit operations are concludable in distinct DTs, which allows optimization, automated process monitoring as well as intelligent control and process state prediction (e.g., early-warning system). In the future, even automated process as well as product validation are feasible with the support of a comprehensive DTMS. Especially, the resulting increased density of information related to the individual assignments to DTs offers the company new possibilities for product and process monitoring, to increase the scope of networking, and for integrating measuring

points that previously had no real-time process connection. Using DTs and the associated peripherals, all relevant information in the value chain, from raw material producer over the bioproduct manufacturer (e.g., food or pharmaceutics) to consumer, can be digitized. Concepts and strategies can be incorporated in the decision-making and quality assurance of bioprocesses.

Approaches from the field of artificial intelligence are a sophisticated method to structure and evaluate the copious data. Sharing information from DTs supplements the DTs with information from other DTs in the B2B sector. Intelligent algorithms can use the information gained from the data for production planning and control. The combination of established simulation and optimization methods as well as the methods from the field of artificial intelligence and machine learning represents a future-oriented technology. Besides, by processing data, KPIs, or forecasts and events information may be included in the optimal configuration of system parameters. Therefore, predictions or behavioral analysis can be applied, and in some cases, repairs of equipment (e.g., fermenters, pumps) can be ordered before the downtime occurs. The digital strategy *Predictive Maintenance* reduces downtimes and associated costs [87, 88]. These predictions are also applicable to resources. Thus, detailed planning of raw material orders can be conducted, or energy can be smartly distributed within the company or, if necessary, diverted to the respective consumer. Furthermore, a detailed production layout planning can be implemented with the help of DTs, where all process resources are used in a time-optimized manner [89]. Hygiene concepts, hygiene state of the equipment, and the hygienic monitoring of associated cleaning concepts can be monitored by DTMS. Moreover, structured DTs in a DTMS help distribute the planning between production units and production systems, improved decision support by simulation models, production unit planning, and automated execution of offers and orders [4].

The comprehensive data collected and evaluated supports documentation of the entire life cycle. This aspect enables higher traceability and new analyses on quality assurance. In particular, the cross-linking, distant from rigid linear hierarchies, creates higher transparency and, thus, new approaches to monitor and analyze processes. A holistic view of the production and the company leads to new means and methods of standardization, which lead to higher efficiency. Plant or product data can be processed directly in the cloud and provided in real-time, and any deviations are detected swiftly. Accordingly, one of the main challenges is to find the root cause of the event within an acceptable time frame and without much effort. Here, machine learning-based anomaly detection can be used. This approach does not directly lead to a root cause but helps identify conspicuous time ranges and production batches. The DTMS and cloud platforms enable patterns to be recognized through various evaluation algorithms and, thus, processes and products can be evaluated in detail and autonomously.

References

1. Forschungsunion, Acatech (2013) Recommendations for implementing the strategic initiative INDUSTRIE 4.0: Final report of the Industrie 4.0 Working Group. https://www.din.de/blob/76902/e8cac883f42bf28536e7e8165993f1fd/recommendations-for-implementing-industry-4-0-data.pdf. Accessed 3 Jun 2020
2. Negri E, Fumagalli L, Macchi M (2017) A review of the roles of digital twin in CPS-based production systems. Procedia Manuf 11:939–948. https://doi.org/10.1016/j.promfg.2017.07.198
3. Kritzinger W, Karner M, Traar G et al (2018) Digital twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine 51:1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474
4. Rosen R, von Wichert G, Lo G et al (2015) About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine 48:567–572. https://doi.org/10.1016/j.ifacol.2015.06.141
5. Liu S (2013) Bioprocess engineering: kinetics, biosystems, sustainability, and reactor design, 1st edn. Elsevier, Amsterdam
6. Taylor DH (2005) Value chain analysis: an approach to supply chain improvement in agri-food chains. Int J Phys Distrib Logist Manag 35:744–761. https://doi.org/10.1108/09600030510634599
7. Fearne A, Garcia Martinez M, Dent B (2012) Dimensions of sustainable value chains: implications for value chain analysis. Supply Chain Manag 17:575–581. https://doi.org/10.1108/13598541211269193
8. Barrientos S (2001) Gender, flexibility and global value chains. IDS Bull 32:83–93. https://doi.org/10.1111/j.1759-5436.2001.mp32003009.x
9. Alfaro L, Antràs P, Chor D et al (2019) Internalizing global value chains: a firm-level analysis. J Polit Econ 127:508–559. https://doi.org/10.1086/700935
10. Trienekens JH (2011) Agricultural value chains in developing countries a framework for analysis. Int Food Agribusiness Manag Rev 14:51–82. https://doi.org/10.22004/ag.econ.103987
11. Bittner K, Spence I (2002) Use case modeling. Safari books online. Addison-Wesley; safari books online, Boston, MA, Sebastopol, CA
12. Nebut C, Fleurey F, Le Traon Y et al (2006) Automatic test generation: a use case driven approach. IIEEE Trans Software Eng 32:140–155. https://doi.org/10.1109/TSE.2006.22
13. Marsch P, Da Silva I, Bulakci O et al (2016) 5G radio access network architecture: design guidelines and key considerations. IEEE Commun Mag 54:24–32. https://doi.org/10.1109/MCOM.2016.1600147CM
14. Ma X, Tao F, Zhang M et al (2019) Digital twin enhanced human-machine interaction in product lifecycle. Procedia CIRP 83:789–793. https://doi.org/10.1016/j.procir.2019.04.330
15. Siepmann D, Graef N (2016) Industrie 4.0 – Grundlagen und Gesamtzusammenhang. In: Roth A (ed) Einführung und Umsetzung von Industrie 4.0: Grundlagen, Vorgehensmodell und Use Cases aus der Praxis. Springer, Berlin, Heidelberg, pp 17–82
16. Deutsches Institut für Normung e.V. (2014) Integration von Unternehmensführungs- und Leitsystemen – Teil 1: Modelle und Terminologie(62264–1:2013)
17. Redelinghuys A, Basson A, Kruger K (2019) A six-layer digital twin architecture for a manufacturing cell. In: Borangiu T, Trentesaux D, Thomas A et al (eds) Service orientation in Holonic and multi-agent manufacturing, vol 803. Springer, Cham, pp 412–423
18. Guo N, Jia C (2017) Interpretation of cyber-physical systems whitepaper (2017). Inform Tech Stand 4:36–47
19. Ibarz A, Barbosa-Cánovas GV (2003) Unit operations in food engineering. Food preservation technology series. CRC Press, Boca Raton
20. Foust AS (1960, reprinted, with corrections, 1962) Principles of unit operations. Wiley, New York

21. International Society of Automation (2010) Batch control part 1: models and terminology (88.00.01-2010)
22. Weyer S, Schmitt M, Ohmer M et al (2015) Towards industry 4.0 – standardization as the crucial challenge for highly modular, multi-vendor production systems. IFAC-PapersOnLine 48:579–584. https://doi.org/10.1016/j.ifacol.2015.06.143
23. Hallscheidt S, Adomeit N, Manske T et al (2019) 1x1 der Normung: Ein praxisorientierter Leitfaden für KMU. https://www.din.de/resource/blob/69886/5bd30d4f89c483b829994f52f57d8ac2/kleines-1x1-der-normung-neu-data.pdf. Accessed 3 Jun 2020
24. Lu Y, Liu C, Wang KI-K et al (2020) Digital twin-driven smart manufacturing: connotation, reference model, applications and research issues. Robot Comput Integr Manuf 61:101837. https://doi.org/10.1016/j.rcim.2019.101837
25. Deutsches Institut für Normung e.V. Normung auf einen Blick. https://www.din.de/resource/blob/249724/2dcba02aa757691193e37aa80a1a6e17/infografik-normung-auf-einen-blick-data.pdf. Accessed 31 Mar 2020 at 07:48pm
26. International Organization for Standardization (under development) Digital twin manufacturing framework: part 1: overview and general principles (23247-1)
27. International Organization for Standardization (under development) Digital twin manufacturing framework: part 2: reference architecture (23247-2)
28. International Organization for Standardization (under development) Digital twin manufacturing framework: part 3: digital representation of physical manufacturing elements (23247-3)
29. International Organization for Standardization (under development) Digital twin manufacturing framework: part 4: information exchange (23247-4)
30. International Organization for Standardization (under development) Automation systems and integration — industrial data: visualization elements of digital twins (24464)
31. International Society of Automation (2001) Batch control part 2: data structures and guidelines for languages (88.00.02-2001)
32. International Society of Automation (2003) Batch control part 3: general and site recipe models and representation (88.00.03-2003)
33. International Society of Automation (2006) Batch control part 4: batch production records (88.00.04-2006)
34. International Society of Automation (2010) Enterprise-control system integration: part 1: models and terminology (95.00.01-2010 (IEC 62264-1 Mod))
35. International Society of Automation (2018) Enterprise-control system integration: part 2: objects and attributes for enterprise-control system integration (95.00.02-2018)
36. International Society of Automation (2013) Enterprise-control system integration: part 3: activity models of manufacturing operations management (95.00.03-2013)
37. International Society of Automation (2018) Enterprise-control system integration: part 4: objects and attributes for manufacturing operations management integration (95.00.04-2018)
38. International Society of Automation (2018) Enterprise-control system integration: part 5: business-to-manufacturing transactions (95.00.05-2018)
39. International Society of Automation. Enterprise-control system integration: part 6: messaging service model (95.00.06-2014)
40. International Society of Automation (2017) Enterprise-control system integration: part 7: alias service model (95.00.07-2017)
41. International Society of Automation (2015) Machine and Unit States: an implementation example of ANSI/ISA-88.00.01(TR88.00.02-2015)
42. International Society of Automation. Procedure automation for continuous process operations: models and terminology (106.00.01)
43. International Society of Automation (2017) Procedure automation for continuous process operations: work processes (106.00.02-2017)
44. International Society of Automation (2008) Using ISA-88 and ISA-95 together (88.95.01-2008)

45. International Society of Automation. Enterprise-control system integration: TR01: master data profile template (95.01-2018)
46. International Society of Automation (2015) Human machine interfaces for process automation systems (101.01-2015)
47. ZVEI – Zentralverband Elektrotechnik- und Elektronikindustrie e. V. (2015) Modulbasierte Produktion in der Prozessindustrie – Auswirkungen auf die Automation im Umfeld von Industrie 4.0: Empfehlungen des AK Modulare Automation zur NE 148 der Namur, Frankfurt am Main
48. International Organization for Standardization (2015) Quality management systems: requirements (9001:2015)
49. International Organization for Standardization (2018) Energy management systems: requirements with guidance for use (50001:2018)
50. International Organization for Standardization (2018) Occupational health and safety management systems: requirements with guidance for use (45001:2018)
51. International Organization for Standardization (2018) Food safety management (22000:2018)
52. International Organization for Standardization (2013) Food safety management systems: requirements for bodies providing audit and certification of food safety management systems (22003:2013)
53. International Organization for Standardization (2007) Traceability in the feed and food chain: general principles and basic requirements for system design and implementation (22005:2007)
54. Deutsches Institut für Normung e.V. (2000) Chargenorientierte Fahrweise: Teil 1: Modelle und Terminologie (61512)
55. International Electrotechnical Commission (1997) Batch control: part 1: models and terminology (61512-1:1997)
56. International Society of Automation (1996) Possible recipe procedure presentation formats (88.0.03-1996)
57. Heidel R, Hoffmeister M, Hankel M et al (2017) Industrie4.0 Basiswissen RAMI4.0: Referenzarchitekturmodell mit Industrie4.0-Komponente, 1st edn. VDE Verlag GmbH; Beuth Verlag GmbH, Berlin, Wien, Zürich
58. Affairs and Energy, Federal Ministry for Economics (2020) Project GAIA-X. https://www.bmwi.de/Redaktion/EN/Publikationen/Digitale-Welt/project-gaia-x.html. Accessed 21 Apr 2020
59. Otto B, Steinbuß S, Teuscher A, Lohmann S (2019) Reference architekture model 2019: version 3.0. https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf. Accessed 30 Mar 2020
60. Tools.ietf.org, Rfcmarkup Version 1.129d On (2020) RFC 3987 – Internationalized Resource Identifiers (IRIs). https://tools.ietf.org/html/rfc3987. Accessed 21 Apr 2020
61. (2020) eCl@ss: home. https://www.eclass.eu/en/index.html. Accessed 21 Apr 2020
62. International Electrotechnical Commission (2017) Standard data element types with associated classification scheme: part 1: definitions – principles and methods (61360-1:2017)
63. The World Wide Web Consortium (2018) XML schema part 2: datatypes second edition. https://www.w3.org/TR/xmlschema-2/. Accessed 21 Apr 2020
64. Wang XV, Wang L (2019) Digital twin-based WEEE recycling, recovery and remanufacturing in the background of industry 4.0. Int J Prod Res 57:3892–3902. https://doi.org/10.1080/00207543.2018.1497819
65. Wagner C, Grothoff J, Epple U et al (2017) The role of the industry 4.0 asset administration shell and the digital twin during the life cycle of a plant: September 12–15, 2017, Limassol, Cyprus: 1–8. https://doi.org/10.1109/ETFA.2017.8247583
66. Otto B, ten Homplel M, Wrobel S (2018) Industrial Data Spaces: Referenzarchitektur für die Digitalisierung der Wirtschaft. In: Neugebauer R (ed) Digitalisierung: Schlüsseltechnologien für Wirtschaft und Gesellschaft, 1st edn. Springer, Berlin, pp 113–133
67. Capiello C, Gal A, Jarke M et al (2020) Data ecosystems: sovereign data exchange among organizations (Dagstuhl Seminar 19391). https://doi.org/10.4230/DAGREP.9.9.66

68. Steinmetz C, Rettberg A, Ribeiro FGC et al (2018) Internet of things ontology for digital twin in cyber physical systems. In: SBESC 2018: 2018 VIII Brazilian symposium on computing systems engineering: proceedings: Salvador, Brazil, 6–9 November 2018. Conference Publishing Services, IEEE Computer Society, Los Alamitos, CA, pp 154–159
69. Jung C, Eitel A, Schwarz R (2014) Enhancing cloud security with context-aware usage control policies. In: Plödereder E, Grunske L, Schneider E et al (eds) Informatik 2014: Big Data – Komplexität meistern: 22–26 September 2014, Stuttgart: proceedings. Gesellschaft für Informatik, Bonn, pp 211–222
70. Bussard L, Neven G, Preiss F-S (2010) Downstream usage control. In: IEEE international symposium on policies for distributed systems and networks (POLICY), 2010: 21–23 July 2010, Fairfax, Virginia, USA; proceedings. IEEE, Piscataway, NJ, pp 22–29
71. Zrenner J, Möller FO, Jung C et al (2019) Usage control architecture options for data sovereignty in business ecosystems. J Enterprise Inform Manag 32:477–495. https://doi.org/10.1108/JEIM-03-2018-0058
72. International Data Spaces e.V. https://www.internationaldataspaces.org. Accessed 23 Jun 2020
73. (2020) DIN SPEC 27070:2020-03, Anforderungen und Referenzarchitektur eines Security Gateways zum Austausch von Industriedaten und Diensten. https://doi.org/10.31030/3139499
74. (2020) HomeIndustrie 4.0 Recht-Testbed. https://legaltestbed.org/en/start/. Accessed 21 Apr 2020
75. unn | UNITED NEWS NETWORK GmbH (2020) Die Bedeutung von IDS für die Europäische Digital-Wirtschaft. https://www.pressebox.de/pressemitteilung/industrial-data-space-e-v/Die-Bedeutung-von-IDS-fuer-die-Europaeische-Digital-Wirtschaft/boxid/993608. Accessed 21 Apr 2020
76. Siemens AG (2020) MindSphere. https://siemens.mindsphere.io. Accessed 28 May 2020
77. Werner R, Beugholt A, Takacs R, Geier D, Becker T, Sollacher R, Mauermann M, Weißenberg N, Roest M, Istaitih J (2020) Standardized digitalization of an existing pudding production by introducing a digital twin management system. International Dairy Magazine
78. Copa Cogeca, CEMA, Fertilizers Europe et al (2018) EU code of conduct on agriculture data sharing: Guidelines for correct use of agriculture data. http://www.fao.org/family-farming/detail/en/c/1127623/. Accessed 28 May 2020
79. Tebbje S, Karthikeyan G, Friesen M et al Entwicklung einer IT-Sicherheitsinfrastruktur für verteilte Automatisierungssysteme: Schlussbericht zu IGF-Vorhaben Nr. 19117 N. https://serwiss.bib.hs-hannover.de/frontdoor/deliver/index/docId/1626/file/Schlussbericht_V31.pdf. Accessed 27 May 2020
80. Singh S, Shehab E, Higgins N et al (2018) Challenges of digital twin in high value manufacturing. In: SAE technical paper series. SAE International400 commonwealth drive, Warrendale, PA, USA
81. Colombo AW, Karnouskos S, Kaynak O et al (2017) Industrial cyberphysical systems: a backbone of the fourth industrial revolution. EEE Ind Electron Mag 11:6–16. https://doi.org/10.1109/MIE.2017.2648857
82. Biffl S, Eckhart M, Lüder A et al (2019) Security and quality in cyber-physical systems engineering: with forewords by Robert M. Lee and Tom Gilb, 1st edn. Springer, Cham
83. Werner R, Beugholt B, Takacs R et al (2020) Digitalisierung einer bestehenden Puddingproduktion – Implementierung eines Managementsystems für digitale Zwillinge. molkerei-industrie: 24–28
84. EIT Food (2020) https://www.eitfood.eu/innovation/projects/the-development-of-organic-supply-chains-that-drive-fair-transparent-and-healthy-options-for-the-consumer-2020. Accessed 21 Apr 2020
85. Über den CCIT (2020) https://www.cit.fraunhofer.de/de/ueber-ccit.html. Accessed 21 Apr 2020
86. Haße H, Li B, Weißenberg N et al (2019) Digital twin for real-time data processing in logistics. Proceedings of the Hamburg international conference of logistics (HICL). https://doi.org/10.15480/882.2462

87. Susto GA, Schirru A, Pampuri S et al (2015) Machine learning for predictive maintenance: a multiple classifier approach. IEEE Trans Ind Inf 11:812–820. https://doi.org/10.1109/TII.2014.2349359
88. Rajesh PK, Manikandan N, Ramshankar CS et al (2019) Digital twin of an automotive brake pad for predictive maintenance. Procedia Comput Sci 165:18–24. https://doi.org/10.1016/j.procs.2020.01.061
89. Uhlemann TH-J, Lehmann C, Steinhilper R (2017) The digital twin: realizing the cyber-physical production system for industry 4.0. Procedia CIRP 61:335–340. https://doi.org/10.1016/j.procir.2016.11.152

Adv Biochem Eng Biotechnol (2021) 177: 167–184
https://doi.org/10.1007/10_2020_157

Published online: 19 February 2021

Digital Twins: A General Overview of the Biopharma Industry

Michelangelo Canzoneri, Alessandro De Luca, and Jakob Harttung

Contents

M. Canzoneri (✉) and A. De Luca
Merck KGaA, Darmstadt, Germany
e-mail: michelangelo.canzoneri@merckgroup.com

J. Harttung
Industrial Affairs, Sanofi, Gentilly, France

Abstract This chapter gives an industry perspective of how digital twins are tangibly translated, implemented, and used in a biopharmaceutical environment. Technical prerequisites and components including data modeling, the lifecycle, and different skills which are required from people to be put together and collaborate efficiently with digital twins are discussed with practical examples which have been implemented in labs and in manufacturing.

Keywords 3D model, Biopharma industry, Connectivity, Contextualizing data, Critical process parameters (CPP), Critical quality attributes (CQA), Data, Digital twin, Healthcare, Holistic, Human factor, Laboratory, Lifecycle, Logistics, Manufacturing, Ontologies, Operations, People, Physical asset, Predictive, Prescriptive, Quality by design (Qbd), Research and development (R&D), Self-driving, Sensors, Simulation, Supply chain, Taxonomies, Virtual Model, VUCA

1 Introduction

A *digital twin* is the result of the convergence of two coexisting systems, the tangible and real system of a living organism or a nonliving physical entity and its virtual replica which is enabled by real-time data and underlying models through the use of digital technologies.

Digital twins have the ability to provide a holistic understanding of the system by building a network of dependencies between real-time data and their underlying meta information.

Through the use of Internet of things (IoT), advanced data analytics, artificial intelligence (e.g., machine learning, deep learning), and models (descriptive, predictive, and prescriptive) the digital twin becomes a living replicate of the physical entity which adapts to real-time information coming from its originator. In the scope of Biopharma, physical entities can range from a biological cell to a complete factory or supply chain. Models should be formalized either as mathematical models or algorithms in order to enable simulations.

So, considering a digital twin to be a real-time connected virtual-physical system where physical reality and virtual models are continuously connected, we will broadly have different levels of twins:

- The most basic digital twin establishes a one-directional relationship between a fixed physical entity and a fixed virtual model providing real-time transmission of physical parameters to the virtual environment for visualization, analysis, and experimental design. Each side of the digital twin system is fixed in the sense that only parameter values change. For instance, a digital twin of a bioreactor would be fixed in the sense that the bioreactor does not change while parameters such as cell density, temperature, pH, etc., would change.
- A more realistic approach in operational usage requires the virtual-physical system to remain operational as changes are made to the structure either of the

physical environment or the virtual environment models with bidirectional exchange of information. Typical changes could include the size of a bioreactor, variations of cell culture, and process parameter controls.
- Finally, a more advanced level provides the ability to simulate situations in the virtual environment that can then be applied to predict evolution in the physical environment and/or direct changes to the physical phenomenon. In the context of Biopharma, digital twins' potential scenarios would range from adapting control strategies of cell fermentation to optimizing supply chain policies such as stock levels of transport preferences. In this case, the digital twin is in a continuous learning state (supervised or unsupervised).

2 Technical Prerequisites and Components of Digital Twins

2.1 Context

So, in a nutshell, a digital twin consists of: The physical or biological product, the virtual manifestation, and the seamless and bidirectional connection between these two elements. Prerequisites and components will increase as we build increasingly sophisticated levels of twins.

The most essential prerequisite of a digital twin is to have an operational goal and defined value. Key examples would be higher yield for a process, lower variability and deviations, or better asset utilization. It is, however, one of the key missed prerequisites in many projects.

For the first level, static physical entity and virtual model with one-directional data connection, the main prerequisites are the following:

- Sensors on the physical system that can measure its state
- Connectivity to establish the link between the physical and virtual environment
- A virtual model of the physical assets (typically a 3D model in engineering related scenarios but could be also more abstract such as cell models)

The second level of twin where we manage lifecycle changes of both physical entity and virtual models requires the additional components:

- An asset framework to manage the relationship between sensors and the virtual model
- Configuration management of the different components

And to complete the list, for the third level of twin where the virtual model takes on a life of its own and starts being used to drive change in the physical entity, we have these additional elements:

- A dynamic virtual model (that can evolve independently from the "real" system)
- Data from executions of the system in variable conditions
- Data modeling and ontologies

There is also one additional prerequisite going across all three levels of twin with increasing complexity:

- People

For each of these prerequisites, we will provide a basic overview, insights on why they are required, and illustrations of common roadblocks to success in industrial environments. Obviously, funding, be it through private means within companies or through public subsidies, is also a prerequisite in most cases but this topic goes beyond the scope of our investigation.

3 Major Prerequisites

3.1 Sensors

The basics of a digital twin being to create a common state between a physical reality and a virtual representation, we need to have sensors that can measure the state of the physical assets and provide that data.

Such sensors exist on most modern equipment and the growth of industrial IoT is generating a lot of innovation making it easier to add additional sensors to existing equipment. There currently exists a wide range of sensors that will capture the state of a physical environment (temperature, pressure, pH, movement, flow, etc.) and we increasingly see so-called software sensors that will analyze basic physical parameters and convert them into higher level measurements [1]. This will typically be the case for measuring biological processes or using computer vision to extract complex information.

The common roadblocks in this area will usually be converting the signal obtained by the sensor into some useful information for the digital twin. Many older sensors will provide an analog signal where a digital signal is required for our twin. Another common case is where sensors are managed by a PLC that no one has the appropriate competencies to use and the sensor information remains locked within the PLC. Critical approaches to avoid these kinds of issues include defining corporate standards for PLC/sensors upon purchase and implementation that ensure such black boxes do not happen.

Beyond sensors, we shall also usually need data from standard Information Technology solutions such as MES (Manufacturing Execution System), LIMS (Laboratory Information Management System), ELN (Electronic Lab Notebook), etc. However, these are generalizable as sensors where we have a human intermediate between the measurement system and the data capture which we would generally try to avoid as we develop more sophisticated levels of digital twins. It should be noted, however, that the sensors and measurement system only need to be as good as the goal given for the twin and in many cases the required sensor quality required for observability of the model can be quite low.

3.2 Connectivity

One thing is to have measurements coming from a sensor, we also need to transmit that data from the physical asset to the virtual environment.

Traditional approaches will include SCADA and wired networks with a historian solution. Many players increasingly use variations of IoT boxes and standard mobile networks such as wi-fi and to make the connection between automation and digital environments while we are seeing mounting maturity of IoT network approaches in different variations converging towards 5G [2].

In parallel, there are a lot of emerging solutions around Industrial IoT platforms from most traditional industrial software players that make this connection between the edge and digital platforms.

Whatever the solutions, usual complexities come from the interconnection of multiple protocols originating in different worlds (automation vs IT vs telecom mainly) and the variable interpretations of standards creating confusion and inconsistencies. Another major pitfall to avoid is in the translation and volumes of data in different environments. The brutal reality of an analog sensor will usually overwhelm digital infrastructures to make the right conversions and simplifications at the right time in the process is critical. The virtual model will normally only require these condensed data sets to be operational.

3.3 Virtual Model of Physical Asset

We need to create a reality of the physical asset as a virtual model. Otherwise, we are no longer doing a digital twin, we are just doing data capture, analysis, and simulation on measurements. In other words, the virtual model needs to be self-sufficient to understand the system without having direct access to the physical reality we are creating a twin for. For instance, traditional Information Technology solutions such as ERP (Enterprise Resource Planning) have captured data on manufacturing operations but only logically without any representation of the actual physical factory and manufacturing equipment.

The representation of the physical asset as a virtual model will often take the form of a 3D model using various CAD standards [3]. These models will under usual conditions be produced by engineering teams designing the physical asset, integrating sub-models from the equipment providers contributing the various components of the physical asset. A limited number of standards exist that are quite interoperable at a basic level of modelization.

In this area, we start to see some major roadblocks appearing. Virtual models will in many cases be considered proprietary Intellectual Property of the providers, both engineering and equipment providers. It is therefore critical to ensure that initial contracts include appropriate availability and use rights of such models. Another more subtle roadblock is that these models are developed with a lack of transparency

or respect of standards that basically make them unusable. Defining beforehand the standards, naming conventions, hierarchies, etc., to follow are important elements of success. The reality of today is that models are developed both by equipment providers and by engineering companies, but nobody actually uses these models beyond the short period of design and construction.

In other areas such as biological systems, 3D models are not the most efficient way to represent the system and other approaches such as computational whole-cell modeling are being used [4].

Having access to an open and programmable model where you can bring in and display the measurements provided by sensors is a last and most important constraint.

3.4 Asset Framework

As we move into more dynamic digital twins, we shall start seeing the need for an additional prerequisite. In our basic solution, we would be manually making the connection between a sensor and the corresponding objects of our virtual model. However, this quickly becomes impractical as the number of sensors increases and we start seeing evolutions of the different assets and models.

The key to overcome this challenge is through the use of an asset framework that manages the overall plant and asset hierarchy and creates an abstraction layer between the physical reality and sensors and the virtual models [5].

An asset framework is thus a hierarchical, contextualized, and digitized model to describe all physical assets (including sensors) in the system and their relationships. In this way, individual sensors and their measurements are not just data points but acquire meaning within the digital twin system and become resilient to changes to physical assets or models. A temperature sensor on a bioreactor remains as such even if we move the bioreactor to another factory or upgrade the bioreactor.

Most solutions for connectivity include asset framework capabilities. The issue is, however, double:

In many industrial environments, these asset frameworks were either not implemented upon commissioning of the equipment or they have not been maintained. Basic operations without digital twin initiatives can usually survive without a good asset framework. Digital Twins without it will not remain operational for long.

There will often be multiple asset frameworks, and these become inconsistent over time if they were not so already from the start. It is not uncommon to find asset frameworks in a historian, the ERP environment, MES, and an IoT platform solution. All at the same time with variable standards.

The most obvious way out of this dead-end is to manage the asset framework not as an element of a transactional system (Historian, ERP, etc.) but as true master data within an MDM solution that will provide a single source of reference for other systems. For most organizations, this will be a serious change.

3.5 Configuration Management

Managing change is another prerequisite for a successful digital twin. It should be obvious by now that a digital twin requires the synchronization of a lot of independently moving parts. While this is of course theoretically possible to do manually it quickly becomes complex and error prone.

Ensuring that changes are appropriately validated with the right workflows, that adequate impact analysis is done beforehand, and that consistent configurations of stable states are managed is a job by itself [5]. It will also typically require a dedicated solution. Some companies have managed to develop this as a dedicated solution but in most cases using a standard solution from the PLM world will be the best answer.

No solution is perfect and not in this case either. Leveraging PLM solutions usually developed in discrete industries has benefits but adapting these to the conditions of a Biopharma industry remains a major challenge.

3.6 Dynamic Model

Up to now, we have remained within the limited ambition of having a virtual model that continuously represents the physical environment. As we move beyond that into an area where the virtual models start to have an autonomous existence and influence on the physical environment, we start to create new prerequisites.

The virtual models need to be able to evolve by itself and influence backwards to the physical world, which essentially means that it needs to contain a model whereby it can:

- Enable simulations of system behavior without input from the physical environment. This will usually for a digital twin include a combination of mechanistic and algorithmic models to cover behaviors of the physical assets and biopharmaceutical processes.
- Drive changes in the physical environment based on measurements and predictive simulations. This will normally encompass a sophisticated process control strategy and model that defines how you can interact with the physical reality.

The dynamic model is the combination of the two. Unsurprisingly this is a much more open area where few solutions exist. Yet, having the depth of equipment and process understanding required to design such models and control strategy is a condition for building a useful digital twin. This is the moment where we start using the digital twin not just for information but to actually influence operational outcomes [6].

There are many potential pitfalls in this step. The standard process control standards in operations today are far from the level required to be useful in a digital twin context. Process characterization, understanding, and control will not be at the

level of formalization to be automatically leveraged by a digital simulation. And on the models for how the system operates, we are often faced with the competing approaches of trying a bottom-up approach, component by component, that fails to capture the system dynamics that actually defines what happens; and an end-to-end heuristic model delivering insights and correlations devoid of any physical causality, amplifying noise in measurements more than modeling the real world.

Success in this area requires a convergence of practices in different disciplines which is still very emerging.

3.7 Data

Many digital twins fail due to a lack of data from previous runs of the target process and equipment. Without enough volumes of data many of the modern machine learning algorithms become unusable. While we have seen success from pure mechanistic modeling in many discrete manufacturing operations (Automotive, Aeronautics,. . .), in Biopharma we seem to still require real-world experiments to provide the context for simulations.

Capturing data is not an issue by itself but there are two challenges:

- Providing experimental data requires several actual experimentations. For new operations, this can only come from development activities. To be noted that some industries have successfully created the ability to do virtual experimentation (a case in point is autonomous driving where most major players have successfully created virtual playgrounds to increase the learning of the driving machine learning model). In Biopharma, based on the complexity of the system we shall most probably have to go in the direction of a mixed model.
- Beyond the measurement, you need to understand what it means and be able to reuse it in different context. This is an overall question of appropriately "contextualizing" data. In many cases data has been captured but without tracking the different changes that were done to the system while doing experiments. Thus, you need to "realign" those datasets to make them comparable. The different capabilities discussed above in point IV. and V. are what helps you achieve this.

Generating data will in many cases be one of the longest parts of a digital twin project or barriers to doing it.

3.8 Data Modeling and Ontologies

Good data modeling is always important for any kind of data analysis project. In cases where we are manipulating complex data with high variability, it is critical to raise the bar even more.

Developing formal ontologies for the key data domains of the twin and leveraging existing standard taxonomies is the current good practice for addressing this area [7].

The field is still emerging, so overcoming the two barriers of finding the right skill sets and building on solid initiatives that will become industry standards is a key challenge. An example of an initiative with such a potential is allotrope.org which is building a complete framework of ontologies, taxonomies, and data modeling for analytical test data.

Failing to do the appropriate effort in the early stages of a twin project might save some time but as the twin model starts to evolve the increasing load of managing changes to data model versions and realigning datasets to enable analysis across experiments quickly becomes unsustainable and may completely break the project.

3.9 People

All projects require people of course but digital twins have specific constraints. Looking at the list of prerequisites, you see that a lot of different competencies are required to be put together and collaborate efficiently. Also, many of these experts will need to work slightly outside their traditional comfort, some because the interdependencies of the digital twin create additional stress on the technologies and domains being applied.

Beyond this, there are also major mindset changes required to be successful. You need people to trust data coming from the twin to change how they operate things. This is less obvious than it might seem in many operational environments. Another change in behavior that a digital twin will require is to move from an approach of "Tests → Hypothesis → Confirmation" to an approach of "Simulation → Hypothesis → Confirmation Test."

4 Typical Lifecycle of a Twin

Building digital twins is complex, as it requires a deep understanding of the physical entity and the synchronization of the virtual and physical side of the twin. That is why digital twins generally go through iterations. If we investigate the case of digital twins for industrial assets (production equipment, factories, supply chain,etc.) the first step will generally start while designing a physical entity using CAD-type tools. A relatively limited additional investment connecting sensor data to this virtual model will then deliver a basic twin. In most cases, however, as the physical entity enters operational life the two components will get out of sync and the digital twin falls apart.

The second phase will then often be the development of a digital twin on an existing physical phenomenon. The focus then becomes much more on the design of a virtual model of this, both the 3D type modeling and more importantly dynamic

modeling of the processes taking place in the physical entity. While in the first approach the focus was on the design and construction of the physical entity here the focus will usually be more on the understanding and optimization of the system with a much stronger business goal. Having the experience of how the physical entity operates and changes also makes it easier to construct the capabilities to manage change in each component of the system. This kind of digital twin can then be enriched to provide more predictive and simulation capabilities but once again it will usually start hitting limitations.

Similar dynamics will exist in digital twins of biological systems but with the added challenge of lower level of control over design and of understanding of the physical entities internal mechanisms.

The ideal lifecycle is at a higher level of maturity where in the initial design equal focus is given to the physical entity and the virtual model as well as their interactions. This also requires designing into the system from the beginning of how the virtual simulations can influence and direct the physical reality. In the case of preexisting physical phenomena such as biology, this of course requires a very deep understanding of that entity in order to formalize upfront what are the characteristics of the physical phenomenon that can be influenced or controlled and embedding in the virtual model a control strategy where outcomes of virtual model simulations can be applied to the physical entity with a continuous feedback loop.

5 Digital Twin: Potential Applications in Healthcare and Biopharma Industry

Digital Twins acting as a digital replica for the physical object, product, and/or service are the next source of competitive advantage for the Healthcare sector.

They are key to accelerate the move towards preventive and personalized medical treatments by modeling reality with advanced analytics techniques in such a way that problems can/will be predicted whether and when they occur, providing the time necessary to treat the patient in advance.

Furthermore, digital twins can provide a safe (virtual) environment for testing the impact of changes on the performance of a specific system or living body. This will enable optimal solutions, as they imply minimal capital investment, but more importantly will drastically limit risks which are obviously critical in the health sector.

A digital twin can be designed and implemented to improve "care-delivery" services, patient experience, and specific treatments as well as the overall Healthcare value chain. The potentials are unlimited as today we are just "looking" at the peak of the iceberg.

However, like in any technology, the misuse of digital twin can drive issues both on the data privacy standpoint as well as accessibility of the potential solution that,

due to the high cost of entrance, may drive inequalities vs patients. Hence, the need for a strong governance body that would regulate the transparency of data usage, data privacy, and type of personalized treatment.

Let us consider what it could mean for a *single individual as a patient*: the pairing of the virtual and physical worlds allows analysis of real-time data and monitoring of responses to treatments, as well as tracking the results of behavior and lifestyle modification, to prevent any type of disease prior to their occurrence. For example, a wearable sensor (iWatch, fitbit, Oura ring, etc.) could track a patient's blood pressure, body temperature and link that information, as well as data on the patient's lifestyle and genetic characteristics, to a digital twin. Consequently, a doctor can develop and test upfront on the virtual twin the most suitable medical plan and lifestyle recommendations that, upon positive confirmation, could then be implemented on the "living" patient.

Another example, within the Healthcare industry, could come from virtualizing a hospital system to create a safe environment in which to test the impact of potential change on system performance. One of the most typical issues that hospitals face is the dilemma between immediate availability of medicines, stock management due to limited space, and the relative lifecycle of drugs that may create "expired medicines." A digital twin can simulate these complex scenarios and identify the right trade-off between inventory on-hold and availability of medicines.

Furthermore, there are many potential use cases for digital twins in *R&D and OPERATIONS* (Manufacturing and Supply Chain) including real-time monitoring, simulation, modeling, and virtual (remote) control of physical assets. Digital twin is playing a key role in manufacturing process development and many operations processes optimization, in harmonizing products with processes and in defining a holistic process control strategy and ultimately realizing the concept of "Quality by design" (QbD). The QbD is a systematic approach in pharmaceutical drug development to ensure predefined product quality by identifying and understanding the impact of all critical process parameters (CPP) on all and critical product quality attributes (CQA). The use of digital twins helps the pharmaceutical industry in aiming for the reduction of drug development and manufacturing cost as well as reducing the timelines for getting drug candidates to the patient.

The fast progress in IoT will certainly accelerate the adoption of this technology. In fact, according to a recent Gartner research, 75% of the companies using IoT in manufacturing will develop a Digital Twin within 12–18 months.

Finally, in the next section, we will make a deeper dive into two real case studies of digital twins implemented in the pharmaceutical and healthcare industry.

6 Digital Twin: Case Study from Merck KGaA Darmstadt, Germany

6.1 Overall Approach

In order to maintain competitiveness in this "new world," Merck KGaA, as a vibrant Science and Technology company, has been significantly investing since several years in digital technology for manufacturing and supply chain to pursue its vision of "Self-Driving Supply Chains."

We envisaged a Healthcare Supply Chain where the DEMAND is "automatically predicted," without any need of human intervention, through a combination of Statistical Forecasting, based on machine learning, and Predictive Forecasting, based on Real World Evidence data thanks to advanced analytics techniques able to integrate structured and unstructured data, in order to derive a forward-looking signal.

Likewise, SUPPLY is "proactively prescribed," through Control Towers (a sort of center of expertise and decision-making for supply chains), based on Digital Twin systems modeled by real-time information flows, enabling production synchronization across the overall E2E network. This enables a LOGISTICS Distribution system personalized, affordable, and agile.

Figure 1 illustrates this concept:

However, the journey towards the vision is made of a series of important steps, one after the other, that enables a progressive development of the three key elements: Technology, Process, and People.

Our recommended "journey" is made of four steps:

- *Step 1 – Integrated Supply Chain*: whereas we operate in a state of "REAL TIME" End-to-End visibility of SC performance/KPI's, gathered through online descriptive dashboarding, all this supporting the IBP (Integrated Business Planning) process towards "One Number" concept.
- *Step 2 – Predictive Supply Chain*: this is the "FORWARD LOOKING" state in which we are able to get a clear demand signal thanks to predictive capabilities (killing Bullwhip effect).

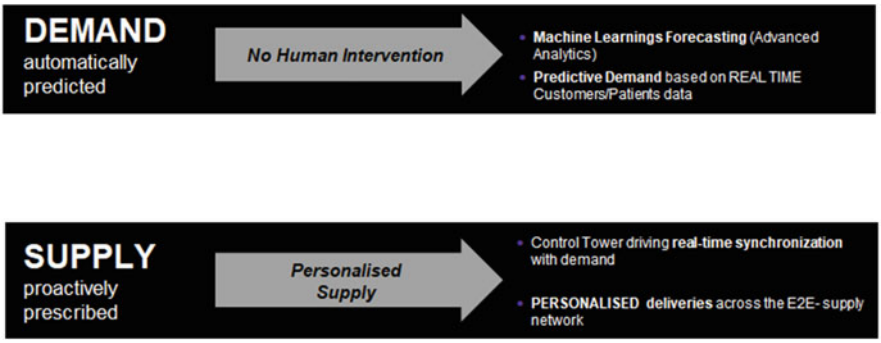


Fig. 1 Self-driving supply chain concept

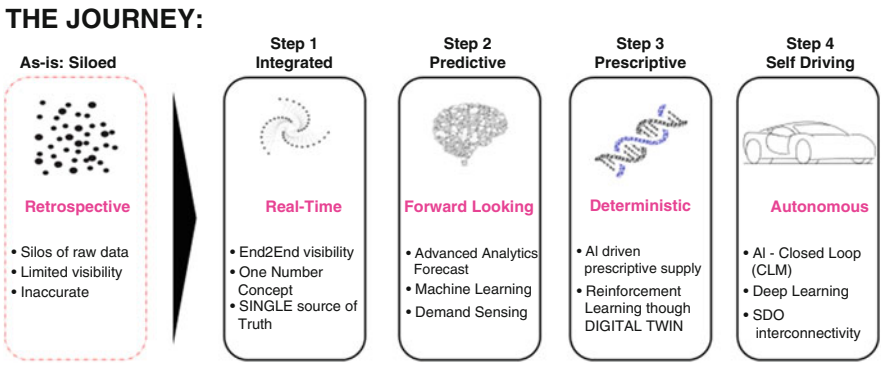


Fig. 2 The journey to implement the vision of a Self-Driving Supply Chain

- *Step 3 – Prescriptive Supply Chain*: moving towards the "DETERMINISTIC" state whereas through the creation of DIGITAL TWINS, the reinforcement learning AI solutions, we have supply planning "prescribed and recommended" by the system (upon human final decision).
- *Step 4 – Self-Driving Supply Chain:* the last step in which the overall balancing of supply vs demand as well as distribution and logistics is totally AUTONOMOUS.

Figure 2 shows the step-by-step approach:

6.2 What Is the Technological Backbone?

Nearly all Pharma companies (and not only) come from a world whereas "ERP is the king." ERP were the systems needed to reduce operational costs, to drive efficiencies through effective data gathering, and to leverage synergies by standardizing processes. While this is (was) generally true, true E2E-Supply Chain-level decision-making was never enabled. Hence the key reason why the end-to-end digital supply chain twin is an important milestone: these new systems do sit "above" any ERP or data gathering source and their analytical layer provides prescriptive insights into the interconnected decisions that are inherent in supply chains (Fig. 3).

At its core is a digital supply chain twin is:

- Connected outside-inside
- Gather and process info real-time (is always "ON")
- Autonomous
- Intelligent

Fundamentally, such a system allows the flexibility for performing simulation scenarios and modeling that can be evaluated without having to necessarily conform to the design constraints of your current supply chain nor impacting the operational activities.

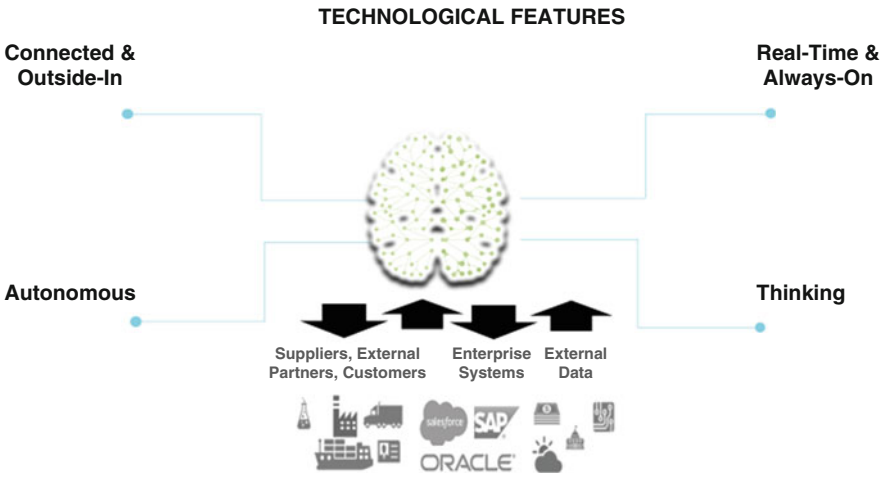


Fig. 3 The technological backbone and its features

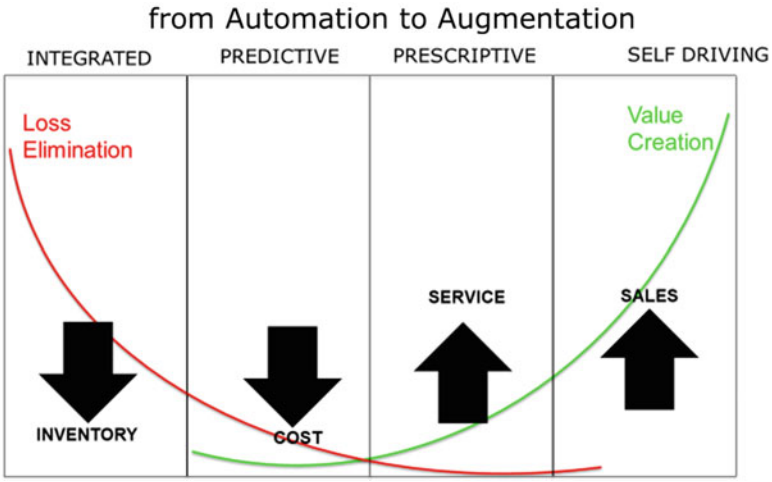


Fig. 4 The journey from Automation towards Automation

It is a journey from "Automation" where companies are increasingly automating processes and eliminating losses towards "Augmentation" enabled by new cognitive automation platforms drive value creation (Fig. 4).

However, this journey is "not exactly a walk in the park" and the two biggest challenges to overcome are:

1. Ensure a very high level of *MASTER DATA ACCURACY*
2. Empower and support a strong *CHANGE MGT PROGRAM*

The first point, Master Data Accuracy, seems trivial but it is not. Indeed, any digital solution becomes useless if data is not correct.

The second point implies the continuous "re-skilling" and "up-skilling" of the supply chain people as new skills and mindset are required.

Effective Change Management is the crossroad between the success and failure of any digital transformation.

We live in a "VUCA" (Volatility, Uncertainty, Complexity, Ambiguity) world where tons of real-time data, better processing power, and more sophisticated analytical algorithms are starting to create seismic shifts in how the supply chain will operate today and in the future.

7 Digital Twin: Case Study from Sanofi

7.1 Objective

This initiative was developed by Sanofi and partially demonstrated at the Usine Extraordinaire event took place in Paris in November 2018.

The objective was to improve operations of a cell culture production line with an endpoint goal of reducing operational incidents leading to lost batches or release delays and to improve overall productivity and yield of production.

7.2 Challenges

The two key challenges are (1) the creation of a model that could reliably simulate the biomanufacturing processes, and (2) ensuring that the operational teams integrate the use of the digital twin in daily activities and are confident with the outcome of the virtual model.

For the first challenge, the main barrier quickly became the availability and quality of data. As we started building models it became very apparent that the data being used to run an industrial process on a daily basis was insufficient to train and validate a model that could reliably replicate the process.

For the second challenge, the issue was different and caused the project to change direction somewhat. While the initial focus was very much on the simulation model, working with the operators it became clear we needed to take a much broader approach where the digital twin was not just a sophisticated technology for a few experts but a solution for the operators in their complete lifecycle experience from training to operations and continuous improvement. Creating operational impact from the digital twin required true operational familiarity and trust from the operational teams.

7.3 Digital Twin Solution

The end approach consisted of (1) building a detailed 3D CAD model of the production line facility and equipment, (2) designing a simulation model for each of the production steps based on historical runs of the same process using key production parameters and measurements and that could predict end-stage values of key outcomes (yield and failed batch) in the initial phases of the step, and (3) connecting sensor data from the process and analytical tests to the model to visualize in context of the equipment.

A first use case was then averaging this model to develop training modules where operators could learn how to operate the process and by themselves use the simulation model to understand and learn how operations parameters and decisions impacted the outcome of the biomanufacturing process.

This training experience then makes it much easier to leverage the twin in operational conditions because the environment and types of decisions are already familiar. Not only does the 3D model provide better training on how to operate an equipment but better understanding of what is happening in the bioprocess increases the right operational procedures.

The second use case was focused on providing shared real-time information on production process status. We confirmed that in many cases the key data on production status is not broadly shared, both because of basic information access and because that data is often managed in technical environments requiring very expert skills to understand its meaning. Providing easily understandable data in the context of the twin is a key enabler in creating the shared understanding of status that avoids operational issues caused by misunderstandings, lack of information, or erroneous data transmission.

The third use case was to enable a set of key decisions on the shop-floor with outcomes of twin predictions. The three main decisions targeted were:

- Decision to stop a batch early that would fail thereby enabling production to save time and quickly restart a new batch
- Decision on timing to end a batch when optimal yield/duration has been hit
- Decision to adjust process control parameters (within the specification of course) in order to optimize yield or avoid deviations and lost batches

7.4 Lessons Learnt

As stated above, integrating the human element of operations was a critical factor in the direction that the twin project took. Creating the visual element of a 3D model and training experiences was a key enabler in operations buy-in and support for the more advanced use cases.

The human factor also drives the need to clearly define and prioritize the operational decisions that the twin is aimed at enabling. It is of course possible to

develop a digital twin for research investigations and broad exploration. However, in an operational environment, the possibilities of a twin can quickly just become information overload and create a lack of trust which is fatal for efficient usage. Therefore, the type of work of proactively defining which decisions are aimed at being enhanced and only focusing on them is a major success factor. If there are too many options, the decisions quickly get evaluated based on the risk one might take versus standard operations without a digital twin instead of the expected operation upside.

Finally, we mainly targeted a modeling approach focusing sequentially on each production step and using combinations of mechanistic and heuristic modeling. While this has the benefits of targeted modeling activities that remain more easily scoped and manageable, it has several caveats. The main issue is probably that as we optimize each step sub-model, we are targeting an endpoint outcome that may actually not be optimal for the end-to-end process. What we optimize for at each step is only as good as our understanding of the overall process.

We explored end-to-end model development using machine learning approaches and while promising it once again becomes apparent that higher volumes of data and better quality are absolutely required for this approach.

8 Conclusion and Outlook

The digital twin approach is still in the emerging phase of its use within the Biopharma industry. We have seen examples of use from process development to manufacturing and supply chain. However, that is only the start and fascinating opportunities also exist in other areas beyond industrial operations. A lot of ink has already been spent around the notions of quantified self and biohacking, investigating possibilities where we take twin approaches to human individuals.

While raising many ethical questions there are also areas such as clinical trials where such approaches would deliver unquestionable benefits. All levels of twins can apply from just having real-time sensors continuously updating a digital replica to the most sophisticated twin where we could do simulations, prediction of therapeutic impact.

The ultimate end goal for digital twin applications within Biopharma and Healthcare would be to no longer have any need for conducting clinical trials on living beings but already a first step would be to shift from an approach of doing experiments and then looking for a model that would explain results to an approach where we execute simulations on the twin to define a predicted best scenario and then verify whether actual therapeutic conditions follow the parameters of the predicted model.

In short, using experiments not to experiment but to validate the hypothesis produced by twin simulations. While radical, one should remember this is already the shift that has taken place in areas such as automotive where nearly all crash tests are now simulated instead of being actually conducted with a prototype vehicle.

In addition, the pairing of the virtual and physical worlds will enable a great step-change towards preventive medicine in a personalized approach. Indeed, the tracking of real-time data, the possibility of analysis and monitoring of responses to treatments in a virtual modeling twin, as well as tracking of the results of behavior and lifestyle modification, will allow the personalized prevention of any type of disease prior to their occurrence. And this is going to be a key breakthrough for science and humanity.

References

1. Chhetri SR, Al Faruque MA (2020) IoT-enabled living digital twin modeling. In: Data-driven modeling of cyber-physical systems using side-channel analysis. Springer, Cham, pp 155–182. https://doi.org/10.1007/978-3-030-37,962-9_8
2. Cheng J et al (2018) Industrial IoT in 5G environment towards smart manufacturing. J Indus Information Integr 10:10–19
3. Tao F, Zhang M (2017) Digital twin shop-floor: a new shop-floor paradigm towards smart manufacturing. IEEE Access 5:20418–20,427. https://doi.org/10.1109/ACCESS.2017.2756069
4. Marucci L, Barberis M, Karr J, Ray O, Race P, Andrade M, Grierson C, Hoffmann SA, Landon S, Rech EL, Rees-Garbutt J, Seabrook R, Shaw W, Woods C (2020) Computer-aided whole-cell design: taking a holistic approach by integrating synthetic with systems biology. Front Bioeng Biotechnol:8. https://doi.org/10.3389/fbioe.2020.00942
5. Weyer S, Meyer T, Ohmer M, Gorecky D, Zühlke D (2016) Future modeling and simulation of CPS-based factories: an example from the automotive industry. IFAC-PapersOnLine 49 (31):97–102
6. Dahmen U, Rossmann J (2018) Experimentable digital twins for a modeling and simulation-based engineering approach. In: 2018 IEEE international systems engineering symposium (ISSE). IEEE, Piscataway, pp 1–8. https://doi.org/10.1109/SysEng.2018.8544383
7. Steinmetz C, Rettberg A, Ribeiro F, Schroeder G, Pereira C (2018) Internet of things ontology for digital twin in cyber physical systems. In: 2018 VIII Brazilian symposium on computing systems engineering (SBESC). IEEE, Piscataway, pp 154–159. https://doi.org/10.1109/SBESC.2018.00030

Adv Biochem Eng Biotechnol (2021) 177: 185–228
https://doi.org/10.1007/10_2020_147

Published online: 23 October 2020

Numerical Methods for the Design and Description of In Vitro Expansion Processes of Human Mesenchymal Stem Cells

Valentin Jossen, Dieter Eibl, and Regine Eibl

Contents

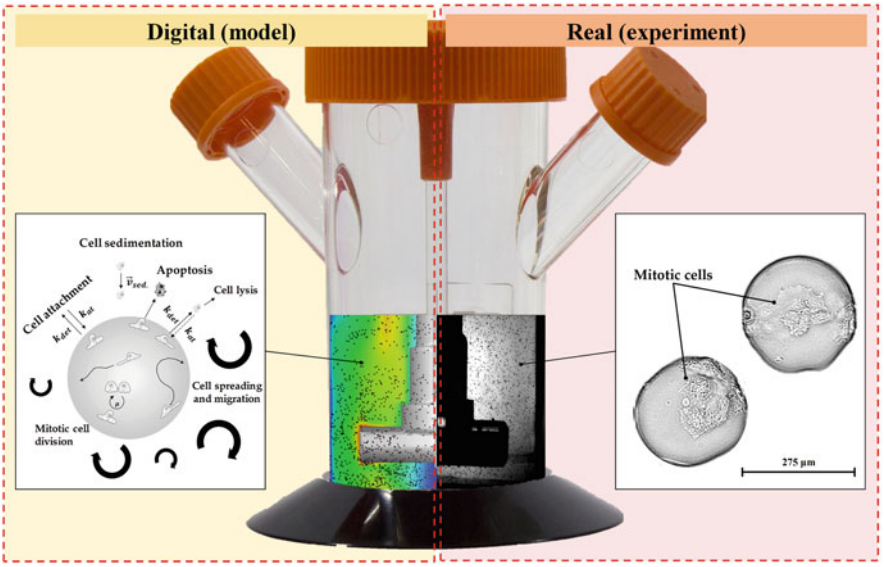
Abstract Human mesenchymal stem cells (hMSCs) are a valuable source of cells for clinical applications (e.g., treatment of acute myocardial infarction or inflammatory diseases), especially in the field of regenerative medicine. However, for autologous (patient-specific) and allogeneic (off-the-shelf) hMSC-based therapies, in vitro expansion is necessary prior to the clinical application in order to achieve the required cell numbers. Safe, reproducible, and economic in vitro expansion of hMSCs for autologous and allogeneic therapies can be problematic because the cell material is restricted and the cells are sensitive to environmental changes. It is beneficial to collect detailed information on the hydrodynamic conditions and cell

V. Jossen (✉), D. Eibl, and R. Eibl
Zurich University of Applied Sciences – Institute of Chemistry and Biotechnology, Wädenswil, Switzerland
e-mail: valentin.jossen@zhaw.ch

growth behavior in a bioreactor system, in order to develop a so called "Digital Twin" of the cultivation system and expansion process. Numerical methods, such as Computational Fluid Dynamics (CFD) which has become widely used in the biotech industry for studying local characteristics within bioreactors or kinetic growth modelling, provide possible solutions for such tasks.

In this review, we will present the current state-of-the-art for the in vitro expansion of hMSCs. Different numerical tools, including numerical fluid flow simulations and cell growth modelling approaches for hMSCs, will be presented. In addition, a case study demonstrating the applicability of CFD and kinetic growth modelling for the development of an microcarrier-based hMSC process will be shown.

Graphical Abstract



Keywords Computational Fluid Dynamics, Euler-Euler model, Euler-Lagrange model, Human mesenchymal stem cells, Kinetic growth modelling, Microcarrier technology, Single-use bioreactor

Abbreviations

CC	Collagen-coated
CFD	Computational Fluid Dynamics
DMEM	Dulbecco's Modified Eagle Medium
DSP	Downstream processing

ECM	Extracellular matrix
bFGF	Basic fibroblast growth factor
FBS	Fetal bovine serum
GMP	Good manufacturing practice
hASC	Human adipose tissue-derived stromal/stem cells
hBM-MSC	Human bone marrow-derived mesenchymal stem cells
hMSCs	Human mesenchymal stem cells
hPL	Human platelet lysate
HGF	Hepatocyte growth factor
HSB	Hemispherical-bottom bioreactor
LDA	Laser Doppler Anemometry
LES	Large Eddy Simulations
αMEM	Modified Eagle Medium
MC	Microcarrier
MCB	Master Cell Bank
MRF	Moving reference frame
OTR	Oxygen transfer rate
PIV	Particle Image Velocimetry
PS	Polystyrene-based
RB	Round-bottom bioreactor
RMSD	Root mean square deviation
SIMPLE	Semi-implicit method for pressure-linked equations
SM	Sliding mesh
SU	Single use
UCM	Umbilical cord-derived mesenchymal stem cells
USP	Upstream processing
VEGF	Vascular endothelial growth factor
VOF	Volume of fluid
WCB	Working Cell Bank

Latin Symbols

Amn (mmol/L)	Ammonium concentration
D_{O2} (m^2/s)	Oxygen diffusivity
D_R (m)	Vessel diameter
EF	Expansion factor
F (N)	Force
Glc (mmol/L)	Glucose concentration
h/H_L	Geometrical ratio between a certain height and the liquid height
h_R/D_R	Geometrical ratio between impeller installation height and the vessel diameter (= off-bottom clearance)
H_L (m)	Liquid height
H_L/D	Geometrical ratio between liquid height and vessel diameter

k_{at} (d^{-1})	Cell attachment constant
k_{det} (d^{-1})	Cell detachment constant
K_{Amn} (mmol/L)	Inhibition constant of ammonium
K_{Glc} (mmol/L)	Monod constant of glucose
K_{Lac} (mmol/L)	Inhibition constant of lactate
Lac (mmol/L)	Lactate concentration
N (rpm)	Impeller speed
N_{s1u} (rpm)	Lower limit of N_{s1} suspension criterion
N_{s1} (rpm)	1s or just suspended criterion (=Njs)
PDL	Population doubling level
P/V (W/m^3)	Specific (volumetric) power input
p_{Amn} (mmol/cell/d)	Specific ammonium production rate (growth-independent)
p_{Lac} (mmol/cell/d)	Specific lactate production rate (growth-independent)
q_{Amn} (mmol/cell/d)	Specific ammonium production rate (growth-dependent)
q_{Glc} (mmol/cell/d)	Specific glucose consumption rate
q_{Lac} (mmol/cell/d)	Specific lactate production rate (growth-dependent)
Re	Reynolds number
r/R	Dimensionless radial coordinates
t_c (s)	Contact time
t_{cir} (s)	Particle circulation times
t_d (d)	Doubling time of cell population
t_l (d)	Lag or cell adaption time
t_{res} (s)	Particle residence time
u_{tip} (m/s)	Impeller tip speed
$\vec{u}$ (m/s)	Velocity vector in x-direction
V_{min} (mL)	Minimal working volume
V_{max} (mL)	Maximum working volume
$\vec{v}$ (m/s)	Velocity vector in y-direction
$\vec{w}$ (m/s)	Velocity vector in z-direction
X_A ($cells/cm^2$)	Cell concentration on surface
X_{max} ($cells/cm^2$)	Maximum cell concentration on surface
X_{Sus} (cells/mL)	Cell concentration in suspension
X_V ($cells/cm^2$)	Cell concentration of viable cells ($X_{Sus} + X_A$)
$Y_{Lac/Glc}$ (mmol/mmol)	Lactate yield per glucose equivalent
$Y_{X/O2}$ (1/mmol)	Yield coefficient/cells per mmol oxygen

Greek Symbols

α	Cell adaption phase coefficient
α_{MC}	MC volume fraction
δ_{Glc}	Step response in glucose balance to avoid negative glucose values ($\delta_{Glc} = 0$ or 1)
η_L (Pa s)	Dynamic viscosity of the liquid

π	Mathematical constant (≈ 3.1415)
ρ_L (kg/m^3)	Density of the liquid
τ_{nn} (Pa)	Local normal stress
τ_{nt} (Pa)	Local shear stress
μ (1/d)	Specific growth rate
μ_{max} (1/d)	Maximum specific growth rate

1 Introduction

The successful development and application of cell-based therapies have the potential to treat a number of currently incurable diseases and to improve patient care. It is therefore not surprising that cell-based therapies have become increasingly important in the field of regenerative medicine, as the expected revenue for 2020 of up to US$ 6.09 billion indicates [1]. Special attention in the field of regenerative medicine is currently being paid to human mesenchymal stem cells (hMSCs). This is unsurprising due to their existence in postnatal tissues (e.g., adipose tissue, bone marrow, the umbilical cord), their high proliferation potential, and their immunosuppressive, immunoregulating, migrating, and trophic properties and low ethical concerns. At the beginning of 2020, 41 clinical trials involving hMSCs were registered (www.clinicaltrials.gov). In addition to the large number of currently ongoing clinical studies, 17 hMSC-based products have received marketing authorization to date (see Table 1), demonstrating the need for reproducible and robust cell processing methods. Product manufacturing takes place mainly with mesenchymal stem cells derived from human bone marrow (hBM-MSC; 11 products), followed by adipose tissue-derived stem cells (hASCs; 5 products).

In general, hMSC-based therapies can be broadly divided into two categories: patient-specific therapies (autologous) and off-the-shelf therapies (allogeneic). From an economic point of view, the allogeneic therapy approach seems to be the most attractive option at present [2, 3]. However, independent of the therapy approach, an in vitro expansion of hMSCs is required to deliver an effective therapeutic dose (1–5 million hMSCs/kg body weight [4–6]). The intention of the in vitro expansion step is to manufacture a sufficient number of hMSCs under good manufacturing practice (GMP) conditions and in a cost-effective manner. It is clear that in vitro manufacturing of hMSCs is often difficult because the cells, which are the product, are directly isolated from body tissue and are genetically unstable in vitro (e.g., cellular senescence) [7]. In addition, significant differences in the cell yield, the proliferation rate, and the differentiation potential have been found between different donors, as well as for different ages of donor and health conditions [8–10]. Apart from the biological variability of the cell material, hMSCs are also sensitive to environmental changes and chemical and physical stresses [11, 12]. As a result, all these aspects place high demands on the in vitro cell expansion process. MSC manufacturing is characterized

Table 1 Available hMSC-based products (as of May 2020)

Medicinal product	Company	Therapy/cell type	Indication	Market
Allostem	AlloSource	Allogeneic ASC	Bone regeneration	USA
Alofisel	TiGenix-Takeda	Allogeneic ASC	Anal fistula in Crohn's disease	EU
AstroStem	Biostar	Autologous ASC	Alzheimer's disease	Japan
aJointStem	Biostar	Autologous ASC	Degenerative arthritis	Japan
Cartistem	Medipost	Allogeneic UCM	Degenerative arthritis	Korea
Cupistem	Anterogen	Allogeneic ASC	Anal fistula in Crohn's disease	Korea
Grafix	Osiris Therapeutics	Allogeneic BM-MSC	Soft tissue defects	USA
HearticellGram-AMI	FCB PharmiCell	Autologous BM-MSC	Acute myocardial infarction	Korea
Neuronata-R	Corestem	Allogeneic BM-MSC	Amyotrophic lateral sclerosis	Korea
OsteoCel	NuVasive	Allogeneic BM-MSC	Spinal bone regeneration	USA
OvationOS	Osiris Therapeutics	Allogeneic BM-MSC	Bone regeneration	USA
Prochymal	Osiris Therapeutics	Allogeneic BM-MSC	Acute graft vs. host disease	Canada
Stemirac	NIPRO Corp	Autologous BM-MSC	Spinal cord injury	Japan
Stempeucel	Stempeutics	Allogeneic BM-MSC	Critical limb ischemia	India
TemCell	JCR Pharm.	Allogeneic BM-MSC	Acute graft vs. host disease	Japan
Trinity Elite	Orthofix	Allogeneic BM-MSC	Bone regeneration	USA
Trinity Evolution	Orthofix	Allogeneic BM-MSC	Bone regeneration	USA

by different manufacturing steps covering upstream processing (USP), downstream processing (DSP), formulation, and fill and finish operations. Typical USP operations are the manufacturing of the Master Cell Bank (MCB) and Working Cell Bank (WCB), seed cell production, and cell expansion at L-scale. DSP operations include cell harvest, cell separation, washing as well as concentration procedures, and medium exchange. Different economic studies have demonstrated that the USP, and in particular the hMSC expansion, represents the main cost driver when examining the whole manufacturing process [3, 13, 14]. To reduce the number of experiments and to increase the process knowledge during either the design and development or the optimization phase, virtual representations of the hMSC production process, so called "Digital Twins," are helpful. These virtual models allow an approximation of real process conditions, a fact that is particularly important for

the production of cell therapeutics, as, among other things, cell material (in an autologous approach) may vary between batches. Process conditions must, therefore, be adapted to the biological starting material, increasing the complexity of the production process. Here application of a "Digital Twin," which combines biochemical engineering data of the cultivation system with a mathematical model of the cell growth, is beneficial, as it tests different process conditions in silico and subsequently proposes optimal parameter combinations for the hMSC production process.

2 In Vitro Expansion Approaches: Current Situation

For the clinical application of hMSCs, the in vitro expansion of the cells represents an important step. Although recent studies have shown the difference in cell yield depending on the hMSC source (e.g., bone marrow vs. adipose tissue), the required therapeutic dose (1–5 million hMSCs/kg body weight) makes in vitro expansion mandatory independent on the hMSC-type. Therefore, different systems and cultivation strategies have been developed over the years for the expansion of hMSCs, which will be presented and discussed in the following sections.

2.1 Planar Approach (2D Cultures)

hMSCs are typically isolated by their capacity to adhere to plastic surfaces. Therefore, the simplest way to expand hMSCs is the usage of plastic vessels, such as T-flask or stacked plate systems, which allow for the expansion of the cells at laboratory and pilot plant production scale for early-phase clinical trials [15]. Planar expansion approaches in normal cell culture flasks (e.g., T-flasks) represent a cost-efficient and easy-to-operate solution. Maximum cell densities for hMSCs from the human bone marrow, the adipose tissue, and the umbilical cord have been reported in the literature in the range of 0.05 to 1.0×10^5 cells/cm^2 (PDL 2.8–7.4) for T-flask cultures performed with serum-containing and serum-free cell culture medium (see Table 2). Maximum cell densities for CellSTACK cultures were even reported in the range of 2.5 to 4.2×10^5 cells/cm^2 (=1.59-2.67 $\times 10^9$ cells) using hMSCs from the bone marrow.

However, scale-up of such an hMSC expansion process would require a large number of cell culture flasks, which is by any means neither economic nor ecologic. Moreover, handling of multiple flasks in parallel is very labor and cost intensive (increased facility footprint) and may result in high flask-to-flask variabilities. In addition, the risk of contamination (e.g., bacteria, mycoplasms) is increased due to the large number of open manipulations. Alternatives to the normal cell culture flasks are stacked-plate or multi-tray culture systems, such as cell factories, which significantly increase the efficiency of the cultivation step by using several layers per cultivation system (up to 40-layer systems available). Thus, the absolute cell number

Table 2 Overview of hMSC expansions in different static, planar cultivation systems

MSC type	2D cultivation system	Culture medium	Cell density	PDL	Ref.
hBM-MSC	T-flask (Greiner)	αMEM + 15 % FBS	0.05-0.6 × 10^5 cells/cm^2	5.6 ± 1.8	[10]
	T-flask (CellBIND)	Corning stemgro hMSC	1.0 × 10^5 cells/cm^2	4-5	[16]
	CellSTACK-5	DMEM/αMEM + hPL	0.4-0.9 × 10^5 cells/cm^2	n/a	[6]
	CellSTACK-10	BD Mosaic SFM	2.5 × 10^5 cells/cm^2	n/a	[17]
	CellSTACK-10	DMEM + 10 % FBS	4.2 × 10^5 cells/cm^2	n/a	[17]
	Nunc Cell Factory-4	αMEM + 10 % FBS	1.8 × 10^5 cells/cm^2	4.9	[18]
hASC	T-flask (Corning)	UrSuppe SFM	0.7 × 10^5 cells/cm^2	2.8-3.2	[19]
UCM	T-flask (Sarstedt)	DMEM + 10 % FCS	0.5 × 10^5 cells/cm^2	4.9	[20]
	CellSTACK-5	DMEM/αMEM + hPL	1.6-1.8 × 10^5 cells/cm^2	n/a	[6]

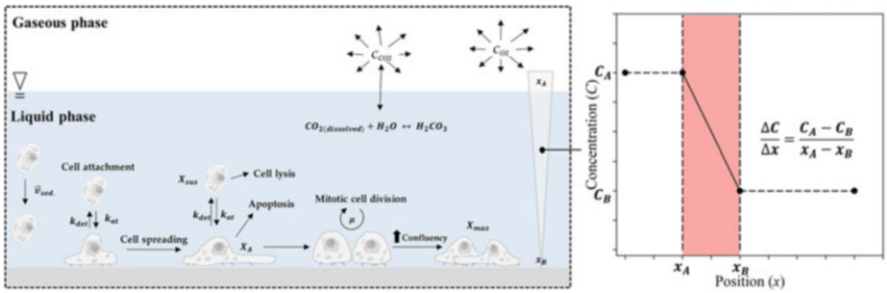


Fig. 1 Schematic representation of biochemical and physical parameters, which have an influence on planar hMSC cultures

per cultivation is significantly increased. Maximum cell densities have been reported in the literature in the range of 0.4 to 4.2 × 10^5 cells/cm^2 for hMSCs expanded in 5- and 10-layer multi-tray systems with serum-containing and serum-free cell culture medium (see Table 2). Due to the static nature of the multi-tray systems, there is always the risk of gradients in pH and pO_2 levels in the liquid phase, possibly introducing heterogeneities that affect cell growth and quality (see Fig. 1). Moreover, the lack of sensors in the systems does not allow the maintenance of optimal set points for some physiochemical parameters (e.g., pH and pO_2), resulting in fluctuating conditions for the cells. The multi-tray systems are also not fully closed, meaning that open manipulations are routinely performed, which require clean room facilities and a class-A laminar flow hood for each manipulation. Interestingly, to date the main reviews on hMSC clinical trials specify that clinical grade cells have mainly been expanded in static 2D systems [6, 15, 21, 22]. However,

in terms of GMP requirements, alternative procedures and cultivation systems, like the spheroid- or microcarrier-based expansion in stirred single-use bioreactors, are said to be the platforms for future cell therapeutic productions (see Sect. 2.2).

2.2 Dynamic Approach (3D Cultures)

As mentioned in Sect. 2.1, hMSCs are typically expanded under adherent conditions as a monolayer in 2D culture systems. However, isolation and growth of hMSCs on rigid tissue culture plastic have been described as promoting spreading of cells rich in actin-myosin stress fibers [23, 24]. Indeed, the static 2D culture systems represent an artificial environment which significantly differs from those of the MSC in vivo niche. Therefore, different efforts have been made over the years to establish dynamic 3D culture systems working with spheroids (see Sect. 2.2.1) or microcarriers (see Sect. 2.2.2). In dynamic bioreactor systems (stirred, wave-mixed, orbitally shaken, hollow fiber and fixed bed types), the culture medium is continuously agitated to provide a uniform environment, preventing the formation of physiochemical gradients and improving mass and heat transfer. Special attention is currently being paid to SU versions, which significantly improve patient safety [25]. Even though different studies have recently shown the applicability of SU systems for MC-based hMSC production processes, challenges still exist.

For this reason, it makes sense to characterize the different bioreactor systems using appropriate process engineering and cell cultivation technique methods prior to usage or during process development, simultaneously assisting in the development of a "Digital Twin." Several studies have been published that provide engineering parameters relating to mixing time, oxygen mass transfer, and power input for various SU bioreactor types. However, when considering the heterogeneous distribution of MCs, spheroids and hydrodynamics, and a detailed analysis of the fluid flow pattern, the MC distribution and the cell growth become worthwhile. Numerical methods, such as Computational Fluid Dynamics (CFD) and kinetic growth models, are complementary methods to the experimental investigations and increase the process knowledge of hMSC production methods. Thus, numerical models can be used to support process development and scale-up.

2.2.1 Growth in Spheroids

hMSCs are often expanded in stirred SU bioreactors as self-assembling cell aggregates or spheroids that mimic the in situ conditions. Thus, compared to 2D monolayer cultures, 3D structures consisting of multiple cell-to-cell contact points are obtained. However, due to their heterogeneous nature, spheroids have been more successfully employed to study complex 3D cell structures and cell differentiation [26] than for hMSC mass expansion in stirred SU bioreactors, as indicated by the limited number of publications in this area (see Table 3).

Table 3 Bioreactors operated with spheroids

MSC type	Bioreactor system	*N*	Medium	Seeding	$D_{max.}$	Ref.
hBM-MSC	100 mL Techne spinner	30 rpm	αMEM+15% FBS	0.2×10^5 cells/mL	135 µm	[27]
	125 mL Shake flask	80 rpm	SFM medium	1×10^5 cells/mL	n/a	[28]
	125 mL Paddle bioreactor	80 rpm	PPRF-msc6	0.5×10^5 cells/mL	218 µm	[29]
hASC	100 mL BellCo spinner	70 rpm	αMEM+10% FBS	6×10^5 cells/mL	350 µm	[30]

The main motivation for growing hMSCs as spheroids is to avoid the use of exogenous support materials, like scaffolds or MCs. Due to the absence of the exogenous support material, the cells are allowed to arrange themselves similar to living tissues [22, 31]. Cells self-assemble and interact under natural forces, permitting them to generate their own extracellular matrix (ECM), which serves as support for the cells to survive in suspension and to mimic the cell-to-cell and cell-to-matrix signaling networks [32, 33]. Investigations by Edmonson et al. [34] have shown that the cell morphology of hMSCs derived from spheroid cultures is comparable to those in bodily tissues. In addition, Caron et al. [35] have demonstrated that a stable hMSCs phenotype is retained in spheroid-based cultures, at least when only the minimum definition of an hMSC is considered [36, 37]. A study by Cheng et al. [38] highlighted that spheroid-derived hASCs exhibited lower cell senescence and a high secretion of angiogenic growth factors (e.g., HGF, VEGF), which was found to be beneficial for wound healing applications. Interestingly, several studies with hBM-MSCs have found that the 3D structure of the spheroids leads to higher yields of secreted immunomodulatory paracrine and anti-inflammatory factors (i.e., TSG-6, stanniocalcin-1, prostaglandin E2) [39, 40], although this was highly dependent on the cell culture medium formulation [41, 42]. The cell culture medium and its formulation play a critical role in spheroid-based hMSC expansions. For example, Zimmermann and McDevitt [41] found that hBM-MSCs expanded in serum-free cell culture medium displayed a reduced expression of prostaglandin E2, indoleamine 2,3-dioxygenase, transforming growth factor-β1, and interleukin-6 when compared with spheroids cultured in serum-containing cell culture medium. Since the cells are forced to aggregate to form spheroids, the medium must also contain adhesive molecules (e.g., laminins, integrins, E-cadherin, vitronectin) to facilitate cell-to-cell attachment [43]. However, for GMP-compliant hMSC productions, these recombinant human proteins represent a strong cost driver, which makes large-scale manufacturing expensive [44]. In addition to biochemical parameters, physical or process engineering parameters have a strong effect on the spheroid culture (see Fig. 2).

For example, oxygen tension has been shown to play a fundamental role in the spheroid formation. Spheroids generated in hypoxic conditions (2% O_2) produced higher amounts of ECM components (i.e., fibronectin, laminin, elastin) and higher

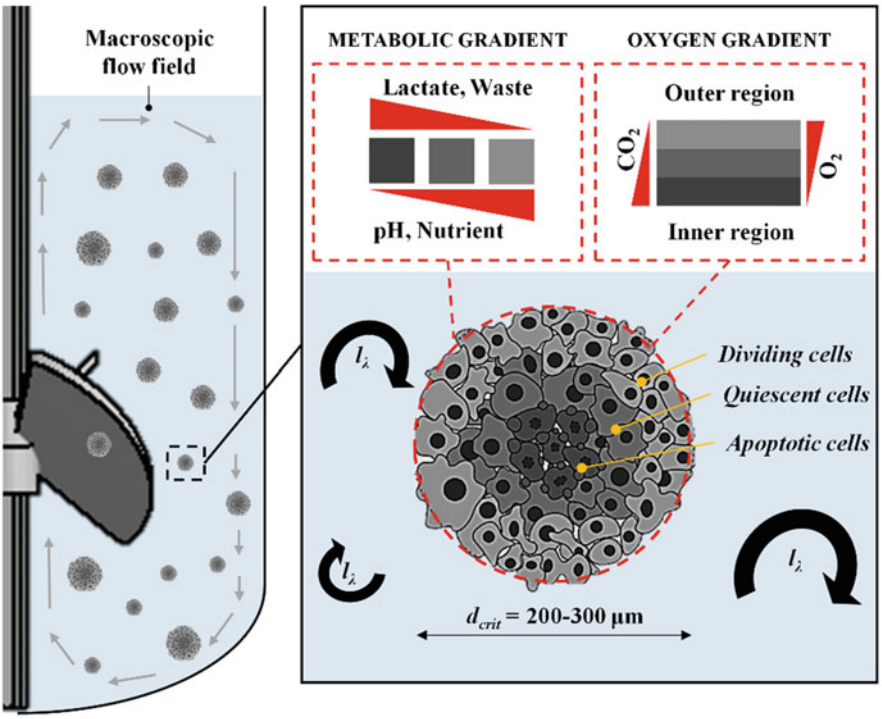


Fig. 2 Schematic representation of biochemical and physical parameters that have an influence on hMSC spheroid cultures

amounts of growth factors (i.e., VEGF, bFGF) [45]. Therefore, spheroids are effective for the tuning of specific cell features but limited in terms of cell proliferation. Bartosh et al. [39] have shown that proliferation-related genes are downregulated in hMSCs upon aggregation. Thus, maximum cell densities in spheroid-based cultures are limited to a certain spheroid size and to the number of spheroids formed in the bioreactor, which limits their applicability for the hMSC mass expansion. Moreover, large spheroids are exposed to diffusional limitations (e.g., oxygen and nutrients), which is a major drawback in high cell density cultures. Different studies have highlighted that spheroids exceeding 200–300 µm tend to induce apoptosis or even undesired spontaneous differentiation due to nutrient or oxygen limitations in the core of the spheroids [46–48]. Indeed, the size of the spheroids can be controlled to a certain level by the fluid flow regime in a stirred bioreactor, but this strategy provides another level of complexity, since spheroid breakage procedures need to be introduced throughout the process. Various studies have shown that the hydrodynamic stresses, the fluid velocities, and the Kolmogorov length scale are very heterogeneously distributed in stirred bioreactors [12, 49, 50], which may limit their effect on the spheroid size. Thus, spheroids are exposed to fluctuating hydrodynamic stresses. Novel bioreactor designs are required that

provide homogenous shear stress levels for the formation and regulation of the spheroid sizes. Such bioreactor development or design studies can be supported by numerical models that allow for optimization of the fluid flow regarding these issues (i.e., homogenous hydrodynamic stress distribution).

2.2.2 Growth on Microcarriers

In order to overcome the limitations of the 2D culture systems, in 1967 van Wezel [51] developed the concept of MC-based cultivation systems. In these systems, the cells are expanded on the surface of small solid particles suspended in the cell culture medium by slow agitation. The MC-based expansion represents a unit operation in which both monolayer and suspension cultures are brought together. The MC surface is available for cell growth, while the mobility of MCs in the medium generates a homogeneity that is similar to the suspension environment used in traditional mammalian submerged cultures [52]. Thus, MC-based expansion systems offer the following advantages:

1. A high surface to volume ratio, which can be further increased by increasing the MC concentration
2. A homogenous environment that allows various process parameters (e.g., pH, pO_2, substrates and metabolites) to be both monitored and controlled
3. A possible scale-up of the MC-based expansion process within a suitable bioreactor series
4. Functionalization of the MC surface to improve cell attachment and in terms of hMSCs to retain a high "stemness"

Different MCs, which are usually spherical, have been tested or even developed over the years for the expansion of hMSCs (see Table 4). The MC types differ greatly in size (90–380 μm), core material (e.g., polystyrene, cellulose, dextran, gelatin), and surface coating (e.g., collagen, fibronectin, laminin, vitronectin). An overview of commercially available MCs, including their material properties, can be found in different reviews [15, 52, 53]. The core material and surface coating affect not only the MC settlement and cell growth but also the impeller speed which is required to hold the MCs in suspension and to guarantee sufficient mass transfer. Rafiq et al. [54] and Leber et al. [55] screened different MC types in small-scale bioreactors for hMSCs under predefined impeller speeds ($N_{js} = N_{s1}$). Both found significant differences in cell attachment, cell growth, glucose consumption, and metabolite production depending on the MC type. They found that hBM-MSC grow best on collagen-coated MCs from Solohill and Synthemax II and ProNectin F MCs from Corning, something which comes as no surprise since these MCs are coated with collagen and fibronectin, respectively. Both coatings are components of the extracellular matrix, including the arginyl-glycyl-aspartic acid sequence which is well-known to promote cell attachment and cell growth of fastidious cells [56]. Different studies have shown that the planar structure, including the material stiffness, nanotopography, and local curvature, can impact cell proliferation, maintenance of

Table 4 Bioreactors operated with microcarriers for the expansion of hMSCs from bone marrow and adipose tissue

MSC source	Bioreactor system	WV	Microcarrier/coating	Culture medium	Agitation	Cell density	Ref.
hBM-MSC	100 mL BellCo spinner	100 mL	Cytodex 1 and 3	DMEM+10% FBS	30 rpm	0.65–0.68 × 10^6 cells/mL	[69]
	100 mL BellCo spinner	100 mL	Polystyrene-based MC	DMEM+10% FBS	30 rpm	0.08 × 10^6 cells/mL (2.7-fold)	[70]
	100 mL BellCo spinner	100 mL	Polystyrene-based MC	PRIME-XV SFM	30 rpm	0.31 × 10^6 cells/mL (10-fold)	[67]
	100 mL BellCo spinner	80 mL	Synthemax II	StemPro MSC	40 rpm	0.36 × 10^6 cells/mL (8-fold)	[68]
	ambr® 15	15 mL	Plastic	DMEM+10% FBS	400 rpm	0.10–0.50 × 10^6 cells/mL	[71]
	BioBLU® 0.3c	250 mL	Cytodex 1 and 3	MSCGM-CD	60 rpm	0.40 and 0.28 × 10^6 cells/mL	[72]
	BioBLU® 0.3c	100 mL	Plastic+PRIME-XV FN	DMEM+10% FBS	115 rpm	0.20 × 10^6 cells/mL	[73]
	BioBLU® 0.3c	100 mL	Plastic+PRIME-XV FN	DMEM+PRIME XV	115 rpm	0.70 × 10^6 cells/mL	[73]
	Mobius® CellReady 3L	2.4 L	Collagen-coated MC	αMEM+10% hPL	25-35 rpm	0.40 × 10^6 cells/mL	[63]
	UniVessel® SU 2L	2 L	CultiSpher G	Lonza medium+5% FBS	70 rpm	0.53 × 10^6 cells/mL	[74]
	BIOSTAT® STR 50L	50 L	CultiSpher G		63 rpm	0.72 × 10^6 cells/mL	[75]
	Mobius® CellReady 50L	50 L	Collagen-coated MC	αMEM+10% hPL	64-100 rpm	0.19 × 10^6 cells/mL	[63]
hTERT-MSC	100 mL Integra spinner	70 mL	Glass-coated MC	DMEM+10% FBS	30-75 rpm	0.38 × 10^6 cells/mL	[55]
	1 L bioreactor Applikon	1 L	Glass-coated MC	DMEM+10% FBS	100 rpm	0.14 × 10^6 cells/mL	[55]

(continued)

Table 4 (continued)

MSC source	Bioreactor system	WV	Microcarrier/coating	Culture medium	Agitation	Cell density	Ref.
hASC	125 mL Corning spinner	100 mL	ProNectin-F	Lonza medium+5% FBS	49 rpm	0.58–1.25 × 10^6 cells/mL	[11, 62]
	100 mL BellCo spinner	80 mL	Synthemax II	StemPro MSC	40 rpm	0.19 × 10^6 cells/mL	[68]
	BioBLU® 5c	3.75 L	Polystyrene-based MC	MSC medium ATCC	25-35 rpm	0.04 × 10^6 cells/mL	[76]
	BioBLU® 5c	3.75 L	Collagen-coated MC		25-35 rpm	0.24 × 10^6 cells/mL	[76]
	UniVessel® SU 2L	2 L	ProNectin-F	Lonza medium+5% FBS	100-140 rpm	0.27 × 10^6 cells/mL	[62]
	BIOSTAT® STR 50L	35 L	ProNectin-F		50-66 rpm	0.31 × 10^6 cells/mL	[62]
hTERT-ASC	125 mL Corning spinner	100 mL	ProNectin-F		49 rpm	0.63 × 10^6 cells/mL	[12]
	500 mL Corning spinner	300 mL	ProNectin-F		52 rpm	0.88 × 10^6 cells/mL	[12]

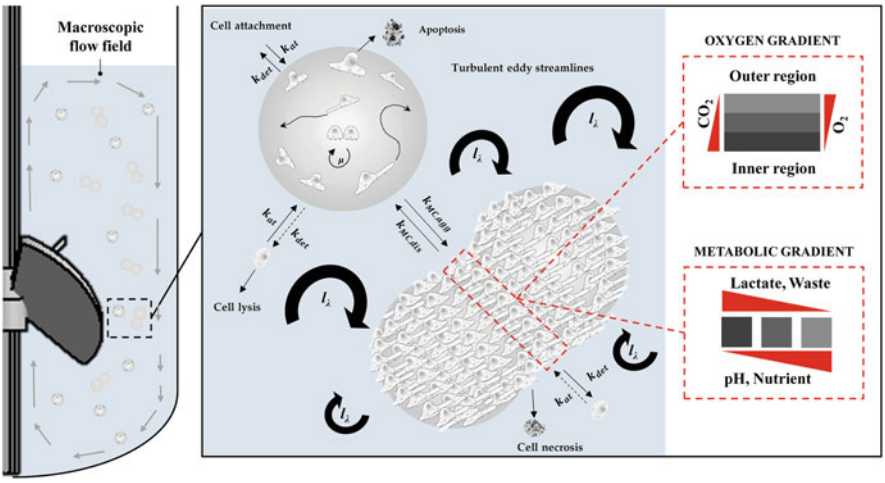


Fig. 3 Schematic representation of biochemical and physical parameters that have an influence on MC-based hMSC cultures

phenotype, and differentiation [57, 58]. Thus, many efforts are being made to develop GMP-grade biodegradable MCs. In general, cell attachment follows a Poisson distribution, where cell-to-MC ratios of one, two, or three result in theoretical probabilities of unoccupied MCs of 0.365, 0.135, and 0.05, respectively [59, 60]. Thus, theoretical cell densities for inoculation are in the range of between 3 and 5 cells per MC. After the cell attachment phase (4–20 h) under static or intermitted stirred conditions, every MC should have the same number of cells attached to its surface. However, in practice, this is not the case. As investigations by Ferrari et al. [61] have shown, suboptimal cell seeding results in the early formation of MC-cell aggregates that impair cell growth and characteristics (see Fig. 3). In addition, large MC-cell aggregates increase the risk of apoptotic cells due to the limited diffusivity of oxygen and nutrients into these aggregates. In fact, the impeller speed can be used to a certain extent to control such MC-cell aggregates, but the hydrodynamic stresses required for this task may also affect the cell growth and quality, especially of the outer cells. To minimize this risk, reliable models of the culture systems ("Digital Twins") are necessary.

In addition to the selection of a suitable MC, the cell culture medium and its formulation also play a key role in the success of a MC-based cultivation. Many of the conventional culture media used for the expansion of hMSCs are defined basal media such as DMEM or α-MEM, which have to be supplemented with additives such as (I) proteins that mediate adhesion to the MC surface, (II) lipids for cellular anabolic purpose, and (III) growth factors and hormones to stimulate cellular proliferation and phenotype maintenance (see Table 4). Even though the disadvantages of serum are well-known, a lot of the hMSC cell culture media additionally contain 5–10% FBS. The highest cell densities generated in serum-containing

medium (10% FBS) have been reported in the range of 0.14–0.65 $\times$ 10^6 cells/mL for cultivations in stirred bioreactors up to benchtop scale. Schirmaier et al. [62] and Lawson et al. [63] reported maximum cell densities of up to 0.3 $\times$ 10^6 cells/mL for cultivations in stirred bioreactors at pilot scale with a cell culture medium supplemented with 10% hPL or 5% FBS. Jossen et al. [11] even reported maximum peak cell densities of up to 1.25 $\times$ 10^6 cells/mL for hMSCs from the adipose tissue in spinner flask cultures with 5% FBS. A proven alternative to FBS is human platelet lysate (5–15%). However, there is still a controversial discussion about whether the cells retain their immunomodulatory properties and their full differentiation capabilities [64–66]. Moreover, there is still a risk of human pathogens and their components being poorly characterized. Therefore, there is a high level of interest in serum- and xeno-free, chemically defined cell culture media. Various formulations are now available on the market (e.g., Mesencult-XF, MSCGM-CD, StemMACS MSC XF, etc.). The careful selection and supplementation of the XF basal medium with suitable growth factors and hormones are important, especially when working with MCs in stirred bioreactors. Special attention has to be paid to cell attachment efficiency and shear stress sensitivity. It is an established fact that the maximum cell densities (0.04–0.40 $\times$ 10^6 cells/mL) and expansion factors that have been achieved in stirred bioreactors with xeno- and serum-free cell culture media are still lower than those achieved in serum-containing medium (see Table 4). Heathman et al. [67] reported a maximum cell density of 0.31 $\times$ 10^6 cells/mL and an expansion factor of 10 within 6 days of using PRIME-XV SF medium in a 100 mL BellCo spinner flask. Carmelo et al. [68] even achieved a maximum cell density of up to 0.36 $\times$ 10^6 cells/mL but a slightly lower maximum expansion factor of 8 with the StemPro MSC medium. Maximum cell densities of between 0.04 and 0.40 $\times$ 10^6 cells/mL were reported for the ATCC and MSCGM-CD medium in the BioBLU 0.3c and BioBLU 5c bioreactor systems.

3 Computational Fluid Dynamics as a Modern Tool for Bioreactor Characterization

Numerical methods, such as CFD, are widely used in the biotech industry to investigate local properties (e.g., flow velocities, shear stresses) in bioreactors and offer an alternative to experimental measurements (e.g., Particle Image Velocimetry (PIV), Laser Doppler Anemometry (LDA)), which are often time-consuming and expensive. Thus, it is unsurprising that CFD is also a valuable tool for the characterization of bioreactor systems used for the production of cell therapeutics. In the following section, a short overview of the basic principle of CFD and various investigations described in the literature are presented. In addition, a case study will be discussed that demonstrates the use of CFD for the characterization of two spinner flask types used for the MC-based hMSC expansion.

3.1 Modelling Approaches

The prediction of the fluid flow is based on solving mass, momentum, and energy conservation equations. This concept includes balances of accumulation, net inflow from convection and diffusion, and volumetric production within an infinitesimally small volume element. For most of the bioprocesses performed in the biotech industry, isothermal conditions (i.e., $T \approx const.$) can be assumed. As a result, the energy balance can be neglected. The mass and momentum equations for incompressible Newtonian media, which includes cell culture media, can be written as shown in Eq. (1) (*Continuity equation*) and Eq. (2) (*Momentum equation*).

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \left(\rho \vec{u}\right) = 0 \tag{1}$$

$$\frac{\partial \left(\rho \vec{u}\right)}{\partial t} + \nabla \cdot \left(\rho \vec{u}\,\vec{u}\right) + \nabla p - \nabla \tau - \rho \vec{g} + \vec{F} = 0 \tag{2}$$

Based on the balancing concept and the spatial discretization of the fluid domain, local and time-dependent data (e.g., velocity gradients, hydrodynamic stress) can be calculated and used for the bioreactor design, the bioreactor characterization, and the process development. Thus, it is unsurprising that different modelling approaches are described in the literature for the CFD-based characterization of bioreactors used for the expansion of hMSCs (see Table 5). For example, Nienow et al. [71, 77], Kaiser et al. [50], Berry et al. [77], and Schirmaier et al. [62] performed single-phase simulations in the ambr 15, the disposable Corning spinner flask, the UniVessel SU 2L, and the BIOSTAT STR 50L based on a *Reynolds-averaged Navier-Stokes* (*RANS*) approach in order to derive the fluid flow pattern and the hydrodynamic stresses acting under different process conditions. The *RANS* approach simplifies the formulation of the instantaneous velocities u by the sum of time-averaged velocities $\overline{u}$ and their fluctuations u', which reduces the computational efforts due to a lower grid resolution. In contrast, Collignon et al. [79] used a *Large Eddy Simulation* (LES) approach, which only resolves macroscopic eddies, for the fluid flow characterization of a 250 mL mini-bioreactor, and their results were found to be in accordance with experimental data. Detailed information about the different numerical models can be found in high-grade textbooks [78–80]. The single-phase simulations do not provide information about the MC distribution and their dynamics in the system. As a result, Delafosse et al. [81], Kaiser et al. [50], and Jossen et al. [11, 12] used a *Euler-Euler* approach in which the MCs were considered as secondary phase. However, this approach does not include discrete formulation of the particle phase and, therefore, only provides information for the entire phase. For this reason, Liovic et al. [82], Jossen et al. [12], and Delafosse et al. [83] described the use of a *Euler-Lagrange* approach which provides a discrete particle formulation and the tracking of individual particles in the bioreactor. Thus, they calculated the circulation and residence times as well as the hydrodynamic stresses acting on individual particles and used this information for process development and characterization.

Table 5 Overview of studies dealing with CFD in order to characterize bioreactor systems for the expansion of hMSCs

Simulation type	Bioreactor system	Title	Ref.
Single-phase (*RANS*)	ambr 15	"The physical characterisation of a microscale parallel bioreactor platform with an industrial CHO cell line expressing an IgG4" and "Agitation conditions for the culture and detachment of hMSCs from microcarriers in multiple bioreactor platforms"	[71, 84]
	125 mL Corning spinner	"Fluid flow and cell proliferation of mesenchymal adipose-derived stem cells in small-scale, stirred, single-use bioreactors"	[50]
	125 mL Corning spinner	"Characterisation of stresses on microcarriers in stirred bioreactor"	[77]
	UniVessel SU 2L and BIOSTAT STR 50L	"Scale-up of adipose tissue-derived mesenchymal stem cell production in stirred single-use bioreactors under low-serum conditions"	[62]
Single-phase (*LES*)	250 mL mini bioreactor	"Large-Eddy Simulations of microcarrier exposure to potentially damaging eddies inside mini-bioreactors"	[85]
Multi-phase (*Euler-Euler*)	125 mL Corning spinner	"Fluid flow and cell proliferation of mesenchymal adipose-derived stem cells in small-scale, stirred, single-use bioreactors"	[50]
	UniVessel SU 2L	"Modification and qualification of a stirred single-use bioreactor for the improved expansion of human mesenchymal stem cells at benchtop scale"	[74]
	1.12 L HSB bioreactor	"Revisiting the determination of hydromechanical stresses encountered by microcarriers in stem cell culture bioreactors"	[81]
Multi-phase (*Euler-Lagrange*)	125/500 mL Corning spinner	"Growth behavior of human adipose tissue-derived stromal/stem cells at small scale: Numerical and experimental investigations"	[12]
	125 mL Corning spinner	"Fluid flow and stresses on microcarriers in spinner flask bioreactors"	[82]
	20L RB bioreactor	"Euler–Lagrange approach to model heterogeneities in stirred tank bioreactors – comparison to experimental flow characterization and particle tracking"	[83]

3.2 Advanced Fluid Flow Characterization of Small-Scale Spinner Flasks: A Case Study

In recent years, various publications in the scientific literature have demonstrated the applicability of stirred SU bioreactors for the in vitro expansion of hMSCs. However, the in vitro expansion processes that provide clinically relevant cell numbers were developed with cell culture media containing 10–20% FBS. The FBS made the

cells more robust and protected against the various stresses (e.g., hydrodynamic stresses, physiochemical stresses, etc.) that occur during the in vitro expansion [86–88]. The focus of this case study is on the biochemical engineering characterization of the Corning spinner flasks (SP100 and SP300) with numerical methods (single- and multi-phase CFD simulations). Special emphasis is placed on the suspension criteria (N_{s1u} and N_{s1}) which are investigated for their use in MC-based hMSC expansions. The case study aims to highlight the use of CFD for the prediction of biochemical engineering parameters and the establishment of a "Digital Twin" to replicate real cultivation systems in silico. For this purpose, multi-phase simulations with a continuum and discrete particle approach were performed, and time-dependent hydrodynamic stresses were derived, based on the transient fluid flow.

3.2.1 Reactor Geometries and Model Approaches

The disposable Corning® spinner flasks (Corning, USA) were commercially available in two different sizes (125 and 500 mL; see Fig. 4). The rigid culture containers were made from polycarbonate and were delivered pre-sterilized. The spinner flasks were equipped with two angled side ports and a 70 mm or 100 mm top cap. The side ports were used for gas exchange (O_2, CO_2) in a standard cell culture incubator.

The main geometrical features of the two spinner flasks are summarized in Table 6. For all numerical investigations, the working volumes were 100 mL (SP100) and 300 mL (SP300), resulting in H_L/D ratios of 0.64 and 0.60, respectively. Both spinner flasks were equipped with a paddle-like impeller consisting of a blade and a magnetic bar. The impellers were directly mounted on the vessel lid and were magnetically driven.

The fluid domain was modelled based on the geometrical data. Subdomains were defined around the impellers in order to implement the impeller rotation using a *Moving Reference Frame* (MRF) or *Sliding Mesh* (SM) approach. In general, unstructured meshes consisting of tetrahedral elements (SP100 = 712,060 CV,

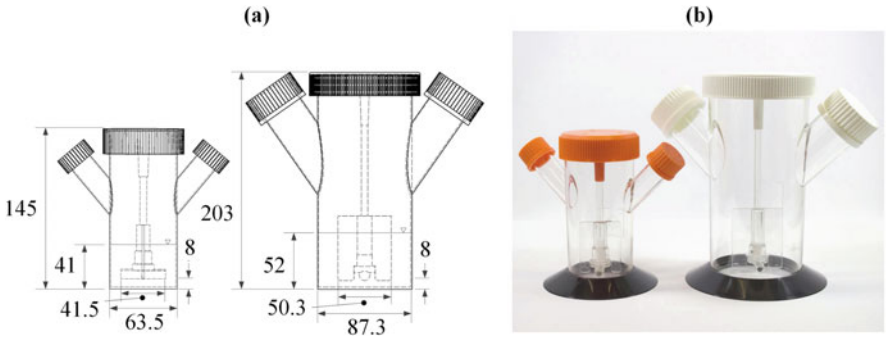


Fig. 4 Small-scale SU Corning spinner flasks (125 and 500 mL) [89]. (**a**) Technical drawings with the main geometrical dimensions (mm). (**b**) Picture of the spinner flasks

Table 6 Overview of main geometrical features of the two Corning spinner flasks

		125 mL Corning spinner (SP100)	500 mL Corning spinner (SP300)
V_{min}	mL	25	50
V_{max}	mL	100	300
D_R	mm	64	87
$H_{L,max}$	mm	41	52
d_R	mm	41	50
h_R	mm	8	8
H_L/D_R	–	0.65	0.60
d_R/D_R	–	0.65	0.58
h_R/D_R	–	0.13	0.09

SP300 = 2,073,079 CV) were used. In addition, a boundary layer along the vessel walls was implemented to improve the resolution of effects close to the vessel walls. The CFD simulations were performed using the *ANSYS Fluent* finite volume solver. The implemented pressure-based solver, with an absolute velocity formulation, was used for all simulations. The walls were treated as non-slip boundaries with standard wall functions. The liquid surfaces were treated as symmetry planes, with the fluid velocities normal to the face set to zero. The MCs were implemented in the simulations using (I) a *Euler-Euler granular* model or (II) a *Euler-Lagrange* approach with discrete particle modelling and tracking. In general, water (ρ_L = 993 kg/m^3, η_L = 0.6913 mPa s at 37°C) and the MC beads ($d_{p,mean}$ = 169 µm, ρ_p = 1,026 kg/m^3) were considered in the models. The initialization of the MCs was carried out either with settled beads (directly at the reactor bottom α_{MC} up to 0.63) or with beads that were homogenously distributed over the entire fluid domain. SIMPLE (semi-implicit method for pressure-linked equations) and phase-coupled SIMPLE algorithms were used for pressure-velocity coupling in the single- and multiphase models. All simulations were run in parallel and solved on a computational cluster (up to 16 Intel Xeno® E5-2630 v4 CPU's @ 2.2 GHz, 64 GB RAM).

3.2.2 Results from Single-Phase Modelling

As shown in Fig. 5a, b, the steady-state fluid flow profiles in the two spinner flask types were similar due to their comparable geometrical ratios. In both cases, the highest fluid velocities occurred at the edges of the impeller blades and in the impeller wake. The maximum fluid velocities were slightly higher (≤5%) than the theoretical u_{tip}, which could mainly be attributed to numerical uncertainties. However, the observations are in agreement with literature data for disk stirrers. For example, Stoots et al. [90] and Wollny [91] demonstrated that the peak tangential velocities in the impeller wake can be up to ≈ 1.4 (experimental) and ≈ 1.5 (numeric) times higher than the impeller speed. An area with relatively weak fluid velocities ($u/u_{tip} < 0.1$) was generated directly below the impeller (r/R ± 0.3) in both systems. Thus, this area represented a critical zone for MC sedimentation. The

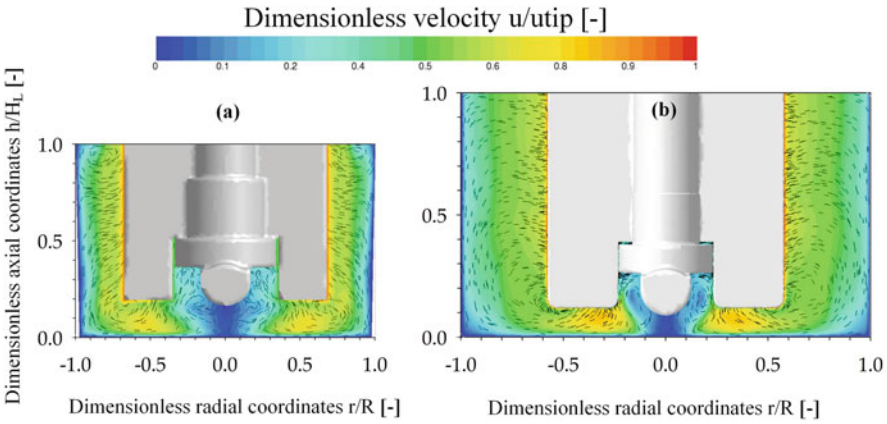


Fig. 5 Steady-state fluid flow inside the SP100 and SP300 [89]. The fluid flow pattern is presented in the vertical mid-plane for N_{s1u}-criterion (SP100 = 49 rpm (**a**), SP300 = 41 rpm (**b**)) as a combined vector and contour plot

observed MC transport from the outer part of the vessel to the vessel center was mainly driven by the induced secondary flow. Similar findings were also reported by Berry et al. [77], Liovic et al. [82], and Venkat et al. [92] in other types of small-scale spinner flasks.

In addition to the stationary fluid flow, the time-dependent behavior of the fluid velocities was simulated for both systems. Compared to the stationary flow field, the occurrence of vortices at the back of the impeller blades becomes visible. According to the definition of turbulence, these vortices occur stochastically and follow the main fluid flow convectively. Similar findings were also reported by Ismadi et al. [93] by means of PIV measurements of small-scale spinner flasks with a slightly different impeller geometry ($d_R/D = 0.88$). The fluctuations in the fluid velocities also become visible when analyzing the fluid velocities at different positions near the impeller (see Fig. 6). It is obvious that after a certain number of stirrer rotations, a "quasi-periodic" fluid movement was obtained. However, the fluctuations in the lower part of the vessel were higher compared to those near the fluid surface. This was not surprising because of the location of the impeller bar which periodically crossed the different areas. Thus, higher fluid velocity gradients occurred in the lower part of the spinner flasks and increased the local turbulences. However, depending on the strength of the velocity gradients, an effect on the cells may be possible. Berry et al. [77] showed that higher fluid velocity fluctuations can result in local hydrodynamic stresses (10^{-3} to 10^{-1} Pa) for the cells in small-scale spinner flasks which are up to three times higher.

Since a number of mathematical assumptions were used for the CFD modelling, stereoscopic PIV measurements were performed to verify the CFD-predicted fluid flow pattern (see Fig. 7). A detailed description of the experimental setup and procedure for stereoscopic PIV measurements can be found in Jossen et al. [12]. For a quantitative comparison of the individual velocity components, the

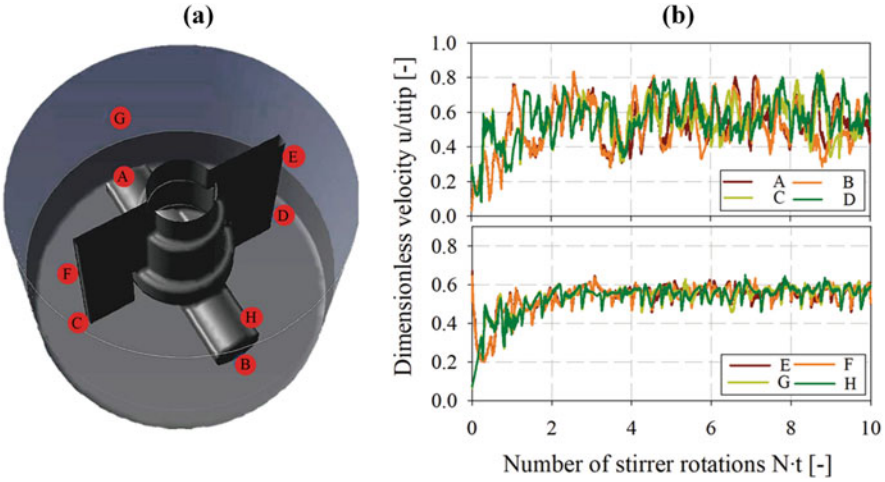


Fig. 6 Time-dependent courses of the fluid velocities at eight different locations within the SP100 [89]. (**a**) Schematic representation of the different locations within the SP100 (= 49 rpm N_{s1u}). (**b**) Dimensionless fluid velocity at the different positions during stirrer rotation

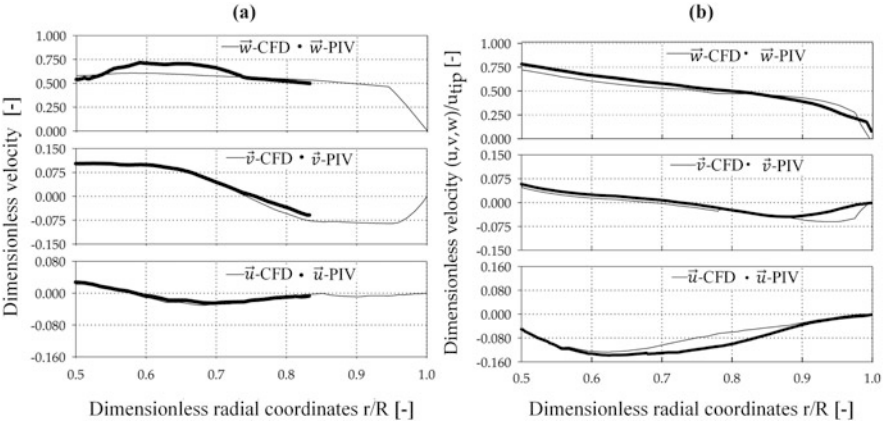


Fig. 7 CFD model verification by experimental PIV measurements in the SP100 and SP300 [89]. Quantitative comparison of CFD-predicted and PIV-measured fluid velocity components ($\vec{u}$, $\vec{v}$, $\vec{w}$) in the SP100 (**a**) and SP300 (**b**)

CFD-predicted and PIV-measured data were compared along dimensionless radial coordinates (0.5–1.0 *r/R*) at an axial position of $h/H_L = 0.1$. The comparison of the velocity components in the SP100 revealed only minor differences for $\vec{v}$ (up to 7.5%) and $\vec{w}$ (up to 8.7%). However, the CFD velocity profiles were well captured, and the overall agreement of PIV and CFD was satisfactory, with findings consistent with those of Kaiser et al. [50]. A comparison of the fluid velocities in the SP100 was

only possible for *r/R* between 0.50 and 0.82 due to the pronounced curve of the vessel surface. The differences between CFD and PIV can be accounted for by measurement uncertainties based on optical phenomena (light refraction and distortion) and the restricted measurement accuracy directly at the edges of the impeller bar (pixel resolution of the camera chip). Thus, direct comparison to the fluid velocities in direct proximity to the impeller is difficult. All three velocity components in the SP300 were well captured by the PIV measurements. The greatest differences (7.9–15%) were found for $\vec{u}$ between r/R 0.70 and 0.85. Hence, it can be concluded that the single-phase CFD model provides reliable fluid flow predictions in both spinner flask types.

3.2.3 Results from Multi-phase Modelling

Oxygen Mass Transfer

Oxygen represents a critical parameter in the cultivation of human cells because it is essential for mitochondrial respiration and oxidative phosphorylation. Hence, the determination of the oxygen mass transfer (*OTR*) represents an important aspect. However, many of the small-scale bioreactor systems frequently used for the expansion of hMSCs are not equipped with oxygen sensors, which makes it impossible to experimentally determine the oxygen transfer. In such cases, multi-phase CFD simulations can be used to estimate the oxygen mass transfer coefficient (k_La), which is shown in the following representative for the SP100.

The multi-phase VOF approach, which takes the headspace into account, was used for the prediction of the k_La in the spinner flasks. Figure 8 (a) shows the stationary fluid flow pattern ($N = 49$ rpm) obtained from the multi-phase VOF model, without significant differences to that derived from the single-phase simulations (see Sect. 3.2.2). This conformity between the single and multi-phase simulations was due to the fact that the transport equations for mass and momentum were corrected only at the phase boundary where both the liquid and the gaseous phase were within the control volume. Since only low impeller speeds ($\leq$120 rpm) were used in the SP100, marginal changes in the fluid surface with relative low interactions between the liquid and gaseous phases occurred. As a result, the multi-phase VOF model also provided reliable predictions for the fluid flow as well as the fluid surface.

The calculation of the k_La value by means of CFD is usually performed in surface-aerated systems using Higbie's penetration model. In this approach, the mass transport is modelled by surface renewal, whereby a characteristic contact time between fluid elements and the phase boundary is calculated (see Eq. (3)).

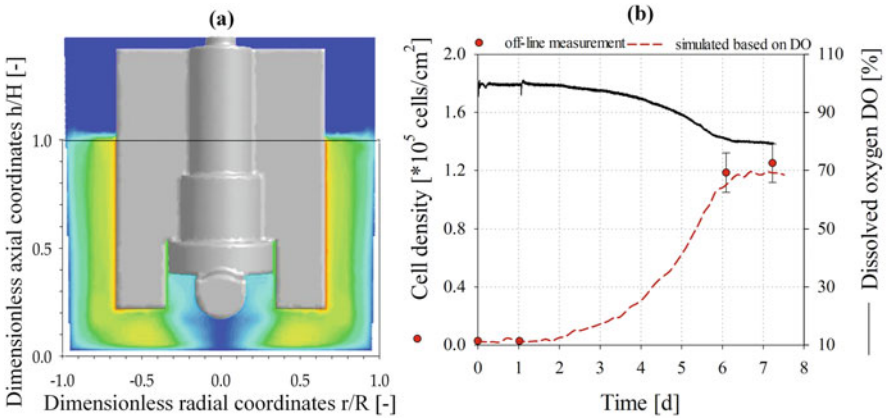


Fig. 8 Fluid flow pattern (**a**) derived from multi-phase CFD simulation and simulated cell growth (q_{O2}) based on data from CFD simulation (**b**)

$$k_L = 2 \cdot \sqrt{\frac{D_{O2}}{\pi \cdot t_c}} \tag{3}$$

Since the fluid flow in the SP100 was mainly tangentially oriented, the contact time was calculated based on the sum of the fluid velocities (w/o the axial component $\vec{v}$) and the mean perimeter of the vessel (see Eq. (4)).

$$t_c = \frac{\pi \cdot d_R}{\sqrt{\vec{u}^2 + \vec{w}^2}} \tag{4}$$

The specific interface area (a) was defined according to Zhang et al. [94] as the area with a liquid volume fraction of $\alpha_L = 0.5$ divided by the total liquid volume (see Eq. (5)).

$$a = \frac{A_{\alpha_L = 0.5}}{V_L} \tag{5}$$

Using this model approach, k_La values of between 2.6 and 4.2 h^{-1} were predicted for impeller speeds between 49 and 120 rpm (= u_{tip} 0.10–0.26 m/s). Compared to experimentally measured k_La values (2.6–4.3 h^{-1}), which were measured in a SP100 specially equipped with an optical pO_2 sensor, only minor differences were found. Consequently, the multi-phase CFD model provided reliable predictions about the oxygen mass transfer in the spinner flasks, especially due to the moderate fluid flow conditions and the surface aeration.

Under consideration of the specific oxygen consumption rate (0.22–2.5 $\times$ 10^{-17} mol/cell/s [89, 95, 96]) or a corresponding yield coefficient for

hMSCs in combination with the oxygen mass transfer, cell growth can be calculated based on the oxygen consumption during the hMSC expansion process (see Eq. (6)).

$$\frac{dX_{MC}}{dt} = k_L a\left(c^*_{O2} - c_{O2}\right) \cdot Y_{X/O2} \tag{6}$$

An example of such an oxygen-dependent growth simulation, which was performed with MATLAB, is shown in Fig. 8b. It is recognizable that the cell density can be simulated based on the current oxygen concentration in the SP100 with a satisfactory accuracy. A good correlation (RMSD = 0.05) was obtained between the simulated and the experimental cell density which was measured offline at the beginning and end of the cultivation.

Microcarrier Distribution Based on a *Euler-Euler Granular* Approach

In MC-based hMSC expansion processes, the sufficient suspension of the MCs is an important aspect since a fully suspended state is desired [96–98]. However, since hMSCs are sensitive to hydrodynamic stresses [99–105], the impeller speed and corresponding power input are limited to a certain level, depending on the MC concentration. Therefore, the characterization of the MC-distribution and the derivation of the acting hydrodynamic stresses are important. One possible numeric approach to obtain these data is the use of a *Euler-Euler granular* model in which the two phases are considered as interpenetrating continua. Therefore, mass and momentum are treated individually for each phase. Figure 9 shows an example of the volume-weighted frequency distribution of the dimensionless MC solid fractions (α/α_{mean}) in the two spinner flasks for a MC solid fraction of 0.1% and for the suspension criterion N_{s1u} (SP100 = 49 rpm, SP300 = 41 rpm). As expected, the highest MC volume fractions were, in both cases, found directly below the impeller in the weak mixing zone ($r/R \pm 0.3$; see also Sect. 3.2.2). This observation is not surprising because of the definition of the N_{s1u}. The spatial position of the CFD-predicted deposits agreed well with those made by Kaiser et al. [50]. They also showed a good correlation of their data with experimental observations, which demonstrates the applicability of the *Euler-Euler granular* model for the prediction of the MC distribution in bioreactors. The CFD-derived volume-weighted frequency distribution of the dimensionless MC volume fractions showed comparable MC homogeneity for the two spinner flask types (see Fig. 9c). The fronting of the distributions clearly indicates zones with low MC volume fractions. These zones were mainly determined near the fluid surface, representing the sedimentation boundary. The similar conditions at the vessel bottom can mainly be explained by the same off-bottom clearance (h_R = 8 mm), whereas the MC distribution over the entire vessel volume is mostly affected by the d_R/D ratio. The results from the two spinner flasks demonstrate that the *Euler-Euler granular* model provides reliable predictions for MC distribution. However, due to the continuum formulation of the

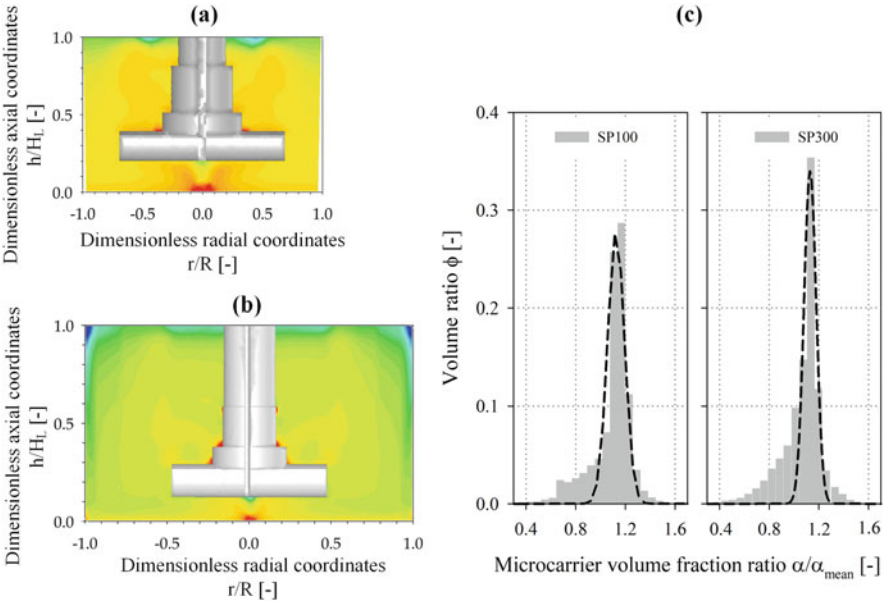


Fig. 9 Contour plots of the dimensionless MC volume fraction (**a**, **b**) and volume-weighted frequency distribution (**c**) at N_{s1u} (SP100 = 49 rpm, SP300 = 41 rpm)

model, information on individual particles and their circulation and residence times in different high shear zones cannot be obtained.

Microcarrier Tracking Based on a *Euler-Lagrange* Approach

Euler-Lagrange simulations allow the spatial distribution of discrete MC particles to be derived. Based on this information, the circulation time ($t_{cir.}$), the residence time ($t_{res.}$), and the hydrodynamic stresses acting on the particles can be calculated. Data from such an *Euler-Lagrange* simulation is shown representatively in the following figure for the SP100. Figure 10a, b shows an example of the fluctuating forces acting on individual MCs during impeller motion. It is obvious that the acting forces fluctuated in the order of 100. Thus, each particle has its own history in terms of hydrodynamic stress, which means that some particles are exposed to a certain hydrodynamic stress level longer and/or more often than others. Compared to the *Euler-Euler granular* approach, which allows volume-weighted data to be derived, the *Euler-Lagrange* approach gives a discrete description per MC.

The particle data can further be processed to derive the force distribution for specific locations or to calculate the circulation and residence times. For this purpose, the two spinner flask types were vertically divided into four zones ($\Delta h/H_L \approx 0.25$). Figure 11 exemplifies the SP100, showing the force distribution in the four defined spinner segments. It is obvious that logarithmic normal distributions

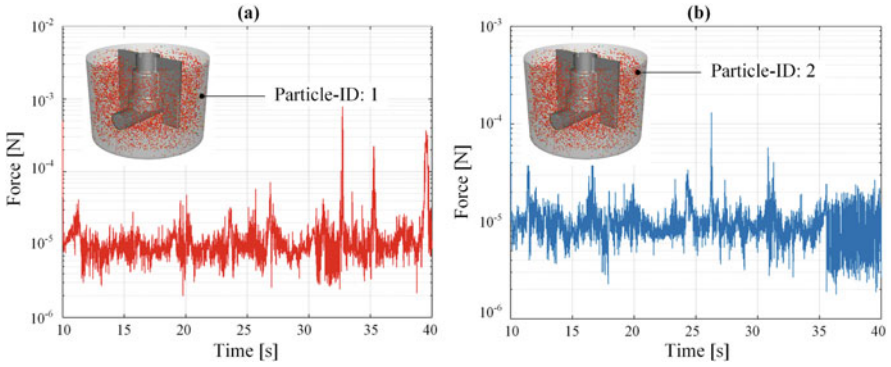


Fig. 10 Force acting on the MCs during the impeller motion. Time-dependent force diagrams are shown representatively for two individual particles in the SP100 ($N = 49$ rpm)

were obtained where highest forces occurred in the lowest segment. Thus, cells on MCs were more stressed in the lowest spinner segment. This observation was also supported by the fact that the highest probability of the presence of MCs was in the lowest spinner segment. However, the effects of the hydrodynamic stresses in the different zones depended heavily on the particle circulation and residence times, demonstrating the dynamics and complexity of the systems. For this reason, circulation times and residence times were calculated for each individual spinner segment based on the particle tracking data and were subsequently averaged over the four segments (see Table 7). As expected, the circulation times (2.7–11.5 s) decreased proportionally to the residence times (0.74–4.94 s) as the impeller speed was increased. Interestingly, the proportionality constants for the SP100 (= 0.54) and the SP300 (=0.49) were quite similar. This observation can be ascribed to the comparable fluid flow conditions. The calculated mean forces were inversely proportional to the circulation and residence times. This finding is not unexpected since the specific power input, which can be calculated based on the torque acting on the impeller during the CFD simulation, increased by approximately the 3rd power in both spinner flask types. Interestingly, the mean values of particle forces did not change significantly between the lower impeller speeds ($N < N_{s1u}$) and the two suspension criteria, even though the circulation and residence times decreased by up to 50%. Impeller speeds exceeding N_{s1u} and N_{s1} resulted in a slight decrease of the circulation times, although the related particle forces increased by exponents of 0.07–0.12 in respect of the resulting specific power input.

Comparable observations for the specific power input are also possible when considering the local normal and shear stresses, which can be calculated according to Wollny [91]. The volume-weighted mean values of the local normal and shear stresses were in a comparable range in both spinner flask types for impeller speeds between N_{s1u} and N_{s1}. Consequently, comparable conditions in terms of hydrodynamic stresses can be expected for cultivations in the resulting specific power input range of 0.3–1.1 W/m^3. Another popular method for evaluating hydrodynamic stress

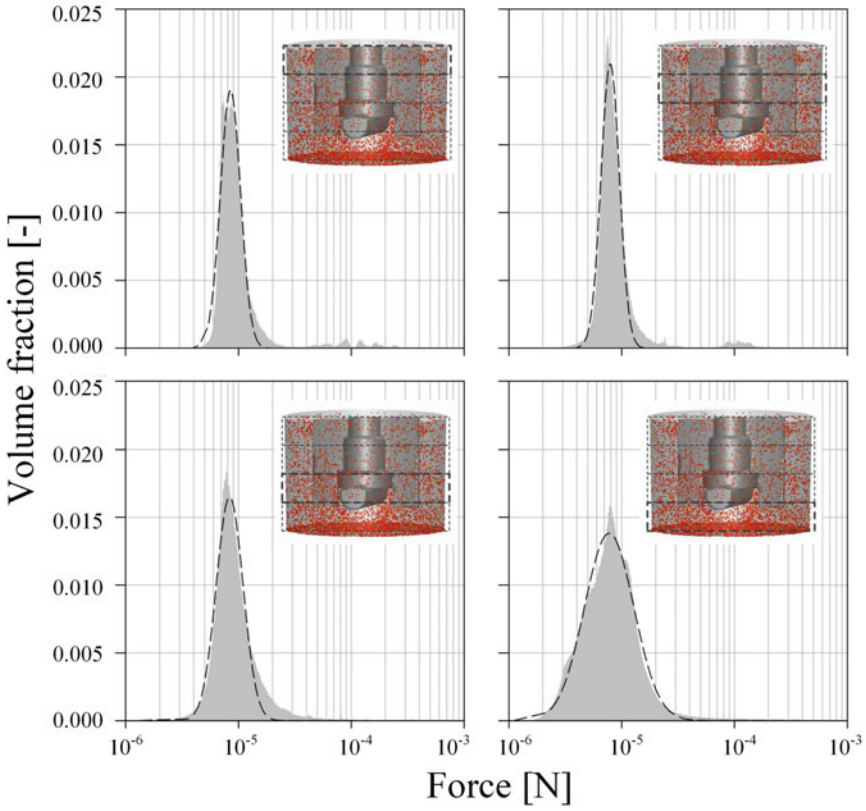


Fig. 11 Force distributions in the different spinner segments

is based on the Kolmogorov length scale, which can be calculated from CFD simulations. While cells in suspension are assumed to only be affected by turbulent eddies of comparable size, those growing on the surface of an MC appear to be more shear sensitive. Croughan et al. [106] found that cell damage became significant when the smallest turbulent eddies were approximately two-thirds of the size of an MC. However, to apply Kolmogorov's theory, the fluid flow must be very turbulent ($Re > 10^4$). The flow in the two-spinner flasks can be described as moderately turbulent. However, the calculated maximum dissipation rates were higher by a factor of two in the impeller swept volume than in the bulk. As expected, the smallest turbulent eddies were found for the highest tested impeller speeds, with values between 30 and 47 μm. In terms of the suspension criteria, the minimum values were predicted between 60 and 76 μm, which is much lower than the proposed two-thirds MC size. In contrast, the volume-weighted mean values were slightly higher than the MC size, which demonstrated that only a small proportion of the turbulent eddies are comparable in size to the MCs. This lowers the risk that the MCs might come into contact with these detrimental eddies. However, this fact also

Table 7 Overview of the main biochemical engineering parameters derived from the CFD simulations

N [rpm]	u_{tip} [m/s]	Re	P/V [W/m^3]	$t_{cir.}$ [s]	$t_{res.}$ [s]	$l_\lambda^{(a)}$ [μm]	$\tau_{nt}^{(b)}$ [10^{-3} Pa]	$\tau_{nn}^{(b)}$ [10^{-3} Pa]	$F^{(c)}$ [10^{-5} N]
Corning 125 mL spinner (SP100)									
25	0.05	715	0.07	11.5	4.9	130/530	2.72/79	0.79/43	0.75
49 N_{s1u}	0.11	1,402	0.63	6.5	2.4	66/228	5.39/169	1.15/108	0.85
60 N_{s1}	0.13	1,717	1.12	6.0	1.9	60/191	6.62/211	1.32/138	0.91
120	0.26	3,434	7.56	4.0	0.9	30/111	12.91/437	2.24/301	1.82
Corning 500 mL spinner (SP300)									
20	0.05	841	0.05	10.0	4.2	136/546	2.04/214	0.30/138	0.83
41 N_{s1u}	0.11	1,724	0.33	6.2	2.6	76/295	4.00/481	0.69/362	0.89
52 N_{s1}	0.14	2,186	0.61	5.9	1.6	66/282	5.00/679	0.87/473	1.04
100	0.26	4,204	3.70	2.7	0.7	47/181	9.26/1,350	1.70/872	2.10

[a]Volume -weighted minimum/mean values of turbulent Kolmogorov length scale
[b]Local shear (τ_{nt}) and normal (τ_{nn}) stress for volume-weighted mean/maximum values
[c]Mean values of acting particle force weighted by number

depends heavily on the resulting circulation and residence times of the MCs. In both cases, the mean volume-weighted values for the highest tested impeller speeds were much closer to the detrimental theoretical value of 141 μm. Even though such eddies occurred at the suspension criteria, the frequency with which the MCs were exposed to such eddies was much lower due to the lower circulation times and residence times.

3.2.4 Linking of CFD-Derived Data with Cultivation Studies

In order to link the CFD-derived engineering data with cell biological aspects, cultivation studies in the two spinner flask types at different impeller speeds were performed. The results of the cultivation studies with hMSCs from the adipose tissue are summarized in Table 8. It is obvious that the different hydrodynamic stress levels have a significant effect on the cell growth in both spinner flask types. Highest living cell densities were achieved, of up to $1.68 \pm 0.36 \times 10^5$ cells/cm^2 ($= 6.25 \pm 0.35 \times 10^5$ cells/mL, EF 56) and $2.46 \pm 0.16 \times 10^5$ cells/cm^2 ($= 8.77 \pm 0.66 \times 10^5$ cells/mL, EF 81), in the SP100 and SP300 when working at $N_{s1u} \leq N \leq N_{s1}$ (SP100 = 49–63 rpm, SP300 = 41–52 rpm). The peak living cell densities in the SP300 were on average up to 40% higher than those in the SP100. Although the two spinner

Table 8 Summary of cultivation results with hMSCs from the adipose tissue in the SP100 and SP300

N [rpm]	*Living* X_{max} [10^5 cells/cm^2]	EF	μ [d^{-1}]	t_d [d]	q_{Glc} [pmol/cell/d]	q_{Lac} [pmol/cell/d]	q_{Amn} [pmol/cell/d]
Corning 125 mL spinner (SP100)							
25	1.05 ± 0.06	35.0	0.6 ± 0.0	1.1 ± 0.1	13.2 ± 2.3	20.7 ± 2.7	8.8 ± 0.3
49 N_{s1u}	1.67 ± 0.12	55.6	0.7 ± 0.0	1.0 ± 0.0	10.6 ± 1.6	35.2 ± 1.9	6.1 ± 0.4
60 N_{s1}	1.68 ± 0.36	56.0	0.7 ± 0.1	0.9 ± 0.1	9.8 ± 0.8	30.3 ± 1.0	6.2 ± 0.3
120	0.60 ± 0.04	20.1	0.5 ± 0.1	1.5 ± 0.4	35.0 ± 1.6	88.8 ± 5.2	16.5 ± 0.3
Corning 500 mL spinner (SP300)							
20	1.36 ± 0.57	45.2	0.5 ± 0.1	1.3 ± 0.1	21.0 ± 0.9	28.6 ± 9.9	14.7 ± 0.2
41 N_{s1u}	2.46 ± 0.16	81.9	0.7 ± 0.0	1.0 ± 0.0	15.5 ± 0.6	40.6 ± 1.8	10.6 ± 0.5
52 N_{s1}	2.43 ± 0.66	81.1	0.7 ± 0.0	1.0 ± 0.0	11.8 ± 1.2	35.3 ± 3.3	9.7 ± 0.4
100	1.25 ± 0.29	41.8	0.5 ± 0.1	1.3 ± 0.0	20.8 ± 9.8	88.6 ± 2.1	19.0 ± 1.4

flask types had comparable geometrical ratios, the hydrodynamic stresses in the SP100 were higher at the suspension criteria. In fact, the absolute hydrodynamic stresses over time were higher due to the lower circulation times, which increase the risk that the cells on the MCs are more frequently exposed to detrimental stresses. At the same time, the residence times, and therefore also the exposure times, of the MCs to the hydrodynamic stresses were shorter, as the multi-phase simulations have indicated. In both cases, the peak cell densities were in the same range as cell densities measured in planar static cultures at maximum confluency ($\approx 2.9 \times 10^5$ cells/cm^2), in which the cells were expanded in parallel. This result indicates that the cells cultivated at $N_{s1u} \leq N \leq N_{s1}$ are mainly restricted by the available growth surface. In contrast, significant lower cell densities were achieved at lower and higher impeller speeds. A peak living cell density of $1.05 \pm 0.06 \times 10^5$ cells/cm^2 ($= 4.49 \pm 0.06 \times 10^5$ cells/mL, EF 35) and $1.36 \pm 0.57 \times 10^5$ cells/cm^2 ($= 4.48 \pm 0.57 \times 10^5$ cells/mL, EF 45) was determined for the SP100 and SP300 at 25 rpm and 20 rpm, respectively. These peak cell densities are up to 84% lower than those at $N_{s1u} \leq N \leq N_{s1}$. This observation may have been caused by the higher amount of sedimented MCs and the increased MC-cell aggregate formation (see also [12]). The viability of the cells on the MCs was always >99%. This was not surprising as dead cells detach from the MC surface. Thus, the increase in dead cells in the supernatant depends on the cell detachment from the MC surface and the die-off of cell in the supernatant.

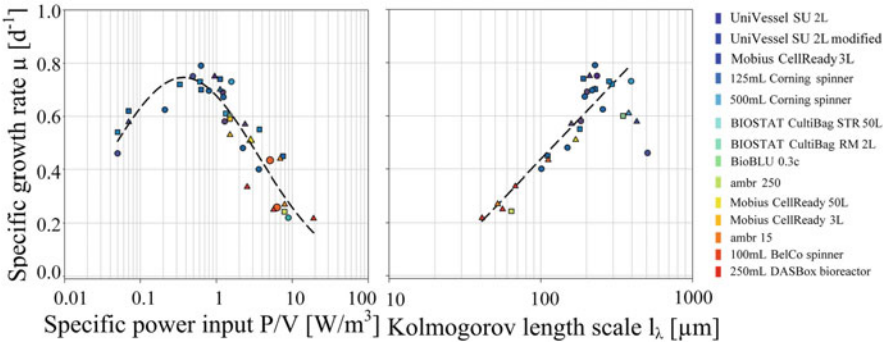


Fig. 12 Dependency of the specific growth rate on the CFD-derived specific power input (**a**) and the Kolmogorov length scale (**b**) [89]. Data from other SU bioreactors were obtained from the literature: UniVessel SU 2L [62, 74], UniVessel SU 2L modified [74], Mobius CellReady 3L [89, 108], BIOSTAT STR 50 L [62, 89], BIOSTAT RM 2L [11], Mobius CellReady 3L [63], ambr 15 [109], 100 mL BellCo spinner [109], 250 mL DASbox bioreactor [73]

By considering q_{Glc}, it becomes clear that the lowest values were obtained for impeller speeds in the range of $N_{s1u} \leq N \leq N_{s1}$ in both cases. This is due to the efficient metabolization of glucose under these hydrodynamic conditions. The calculated values for the hMSCs correspond to those determined by Rafiq et al. [54] and Heathmann et al. [107] in different cell culture media. The highest q_{Glc} (21–35 pmol/cell/d) were found at the highest impeller speeds. The relationship between the q_{Glc} and the specific power input can be expressed by a statistical, logarithmic function of 3rd order. Similar correlations were also found for q_{Lac} and q_{Amn}. However, such statistical correlations are only valid for the investigated *P/V* range. Values of up to 193% and 170% higher than those in the spinner flasks at N_{s1u} and N_{s1} were determined for q_{Lac} and q_{Amn} at the highest impeller speeds. These higher values indicated that the cells are more stressed at higher impeller speeds as a result of the higher hydrodynamic stresses. The different correlations obtained were used as initial parameters for the cell growth modelling (see Sect. 4.2).

Figure 12a, b shows the relationship between the overall mean specific growth rate and the specific power input and Kolmogorov length scale, respectively. The parabolic curve profile of the specific growth rate shows optimal cell growth for $N_{s1u} \leq N \leq N_{s1}$. For specific power inputs between 0.33 and 1.12 W/m^3, maximum μ between 0.70 and 0.74 d^{-1} were achieved. This function also correlates well with literature data from other SU bioreactors. Similar relationships to the specific power input were also established for the Kolmogorov length scale, where a linear relation was found. Thus, CFD-derived hydrodynamic stress data can be used to find correlations between biochemical engineering and cell cultivation aspects and to define optimum cultivation conditions for MC-based hMSC expansion processes.

4 Mathematical Growth Modelling of MC-Based hMSC Expansions

The development of mathematical growth models to describe or predict hMSC growth is gaining in importance. This is not surprising since the cell material is often limited and isolated directly from the patient. Thus, the prediction of the cell growth depending on patient data (e.g., age, health status) is an important aspect, especially for autologous therapies. The following section gives a brief overview of different growth models described in the literature for the expansion of hMSCs. In addition, a case study is presented and discussed, which presents an unstructured, segregated growth model for the expansion of hMSCs on MCs.

4.1 Modelling Approaches

Table 9 gives an overview of publications describing different model approaches for the simulation of the hMSC growth. For example, Higuera et al. [110], Dos Santos et al. [111], and Jossen et al. [12] used kinetic growth models based on Monod-type kinetics. Higuera et al. focused in its formulation only on the substrate/metabolite inhibition, whereas Dos Santos and Jossen et al. introduced terms that considered cell contact inhibition. All models allowed the hMSC cell growth and substrate

Table 9 Overview of hMSC growth models described in the literature

Model type	Title	Ref.
Monod-type kinetic models	"Quantifying in vitro growth and metabolism kinetics of human mesenchymal stem cells using a mathematical model"	[110]
	"Ex-vivo expansion of human mesenchymal stem cells: a more effective cell proliferation kinetics and metabolism under hypoxia"	[111]
	"Growth behavior of human adipose tissue-derived stromal/stem cells at small scale: numerical and experimental investigations"	[12]
Population balance models	"Population balance modelling of stem cell culture in 3D suspension bioreactors"	[112]
	"Experimental analysis and modelling of bone marrow mesenchymal stem cells proliferation"	[113]
	"A mathematical framework to study the effects of growth factor influences on fracture healing"	[114]
	"Modelling of in vitro mesenchymal stem cell cultivation, chondrogenesis and osteogenesis"	[115]
Cellular automaton models	"Population dynamics of mesenchymal stromal cells during culture expansion"	[116]
	"Expansion of adipose mesenchymal stromal cells is affected by human platelet lysate and plating density"	[117]
Cell-based podia model	"Spatial organization of mesenchymal stem cells in vitro – results from a new individual cell-based model with podia"	[118]

consumption to be described based on the experimental setup investigated. In contrast to the Monod-type models, Bartolini et al. [112], Mancuso et al. [113], Bailon-Plaza et al. [114], and Geris et al. [115] used population balance models. For example, Bailon-Plaza et al. [114] included different cell populations in their model in order to describe not only hMSC proliferation but also chondrogenic and osteogenic differentiation. However, all models included parameters strongly influenced by various biological aspects. A discrete formulation of the cells was given by Schellenberg et al. [116] and Cholewa et al. [117], who both used cellular automaton models to describe the hMSC cell growth. However, these models did not include a metabolic description of substrate consumption and metabolite production, which can have an inhibitory effect on the cell growth. Hoffmann et al. [118] developed an individual cell-based model with podia, which is able to quantitatively describe the spatio-temporal organization of MSC culture. They modelled discrete cells and considered their orientation on a planar surface. Hence, the model considers the effects of contact inhibition and the organization and orientation of the cell monolayer. However, the model does also not reflect the metabolization of different substrates or the production of inhibitory metabolites.

4.2 Kinetic Growth Model for the MC-Based hMSC Expansion: A Case Study

Based on theoretical considerations, an unstructured, segregated, simplistic growth model was developed for the MC-based hMSC expansion in the SP100 and SP300. Theoretically, the entire expansion process can be divided into four steps: (I) cell sedimentation and initial attachment, (II) cell spreading and migration, (III) mitotic cell division, and (IV) cell growth arrest due to contact or substrate inhibition, which partially ran in parallel. The general concept of the growth model and the factors that influence the MC-based culture are shown in Fig. 13. During the cultivation period, the formation of MC-cell aggregates is promoted due to the increasing number of cells per bead and periodic particle interactions. The rate of the MC-cell aggregate formation is influenced by the frequency and strength of the hydrodynamic stresses. However, the rate of MC-cell aggregate formation was not considered in the current version of the MC-based growth model because the aggregation process is very complex and depends on many physical and biological parameters. Due to the fact that hMSC growth is anchorage-dependent, possible formation of spheroids in the suspension was not considered in the model. This simplification was justified since no spheroid formation was observed in the MC-based expansions. Thus, it can be assumed that cells in suspension do not contribute to an increase in the overall cell number, with cell growth restricted to the MC surface. To define the starting conditions, it was assumed that initial cell attachment took place during the cell attachment phase, which can be described by the attachment constant k_{at}. After the cells had attached themselves to the MC surface, a short cell adaption phase was

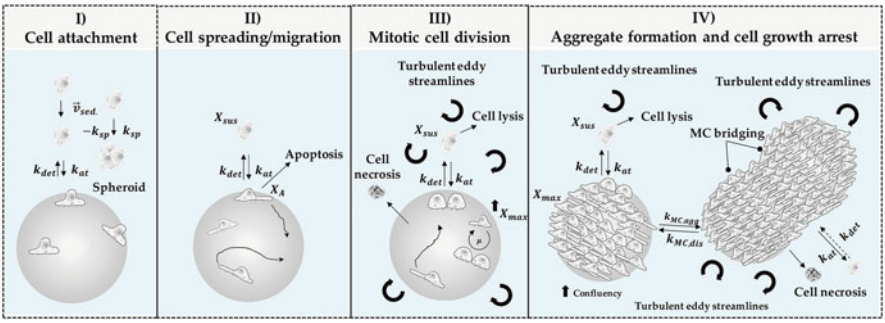


Fig. 13 Schematic representation of different phases and influencing factors during the MC-based expansion of hMSCs. The MC-based expansion can be divided into four phases: (I) cell sedimentation/attachment, (II) cell spreading/migration, (III) mitotic cell division, (IV) MC-cell aggregate formation and cell growth arrest, with some running in parallel

considered before the cells began to proliferate. The cell adaption phase was considered by introducing the coefficient α (see Eq. (7)),

$$\alpha(t) = \frac{t^n}{t_l^n + t^n} \tag{7}$$

where t_l defined the lag time or adaption time and the point at which $\alpha(t)$ is half of the maximum. The exponent n affects the slope of $f(\alpha(t))$. If $n = 1$, $\alpha(t)$ is described by Michaelis-Menten kinetics. Otherwise, a sigmoidal curve is obtained that becomes steeper as n increases. Both variables can be obtained from experimental growth studies.

The specific cell growth rate (μ) was calculated based on Monod-type kinetics. Hence, glucose (Glc), lactate (Lac), ammonium (Amn), and the available growth surface (X_{max}) were considered to be influencing factors (see Eq. (8)). However, investigations indicated that cell growth restriction based on maximum available growth surface does not follow a normal Monod-type kinetic. This fact can mainly be ascribed to cell migration during cell growth. Thus, the effect of the growth surface restriction term becomes more significant towards the end of the cell growth phase. For this reason, the exponent n was also introduced in Eq. (8).

$$\mu = \mu_{max} \cdot \left(\frac{Glc}{K_{Glc} + Glc}\right) \cdot \left(\frac{K_{Lac}}{K_{Lac} + \mathrm{Lac}}\right) \cdot \left(\frac{K_{Amn}}{K_{Amn} + Amn}\right) \cdot \left(\frac{{X_{max}}^n - {X_A}^n}{{X_{max}}^n}\right) \tag{8}$$

The cell number on the MC surface (X_A) increased through mitotic cell division and the attachment of cells from the suspension (see Eq. (9)). However, this increase in cell number was affected by the detachment of hMSCs from the planar growth surface, which was accounted for by the detachment constant ($-k_{det}$).

$$\frac{dX_A}{dt} = \alpha \cdot \mu \cdot X_A + k_{at} \cdot \frac{({X_{max}}^n - {X_A}^n)}{{X_{max}}^n} \cdot X_{Sus} - k_{det} \cdot X_A \tag{9}$$

However, the detachment constant $-k_{det}$ is strongly affected by hydrodynamic forces and is therefore variable for different specific power inputs. As mentioned previously, cell growth in the suspension is negligible, and, therefore, changes in cell concentration will only be affected by attachment to or detachment from the MC surface (see Eq. (10)).

$$\frac{dX_{Sus}}{dt} = k_{det} \cdot X_A - k_{at} \cdot \frac{({X_{max}}^n - {X_A}^n)}{{X_{max}}^n} \cdot X_{Sus} \tag{10}$$

Contrary to the growth restriction based on the specific growth rate, glucose consumption was only limited by the glucose concentration itself (see Eq. (11)). Consequently, glucose consumption was the result of the glucose uptake by the mitotic cells and the maintenance metabolism of mitotic and non-mitotic cells (X_V). A step response (δ_{Glc}) was implemented in Eq. (11) to avoid negative glucose concentrations.

$$\frac{dGlc}{dt} = -\frac{1}{Y_{\frac{X}{Glc}}} \cdot \alpha \cdot \mu \cdot \frac{({X_{max}}^n - {X_A}^n)}{{X_{max}}^n} \cdot X_A - m_{Glc} \cdot \delta_{Glc} \cdot X_V \tag{11}$$

L-glutamine (*Gln*) consumption was not considered in this model since metabolic measurements from the experiment indicated that *Gln* is not a limiting factor. Moreover, UltraGlutamine (L-alanyl-L-glutamine) is used in most stem cell culture medium for which the model was developed and had undergone a series of complex degradation steps (i.e., (I) cleavage by extracellular peptidases and (II) degradation of free L-glutamine or absorption into the cells and metabolization). The production of lactate (*Lac*) and ammonium (*Amn*) was accounted for by Eqs. (12) and (13).

$$\frac{dLac}{dt} = q_{Lac} \cdot X_A \cdot \alpha + p_{Lac} \cdot X_V \tag{12}$$

$$\frac{dAmn}{dt} = q_{Amn} \cdot X_A \cdot \alpha + p_{Amn} \cdot X_V \tag{13}$$

The validity of the unstructured, segregated growth model was tested for MC-based hMSC expansions in the SP100 and SP300 (each $n = 3$), which were performed at N_{s1u} (SP100 = 49 rpm, SP300 = 41 rpm). All growth-related simulations were performed with MATLAB 2019b (MathWorks Inc.) where the model equations were solved using the *ode15s* solver (Intel Core i-7 CPU @ 2.6 GHz, 32 GB RAM). Table 10 shows the parameters and the initial values for the growth simulations which were derived from experimental cultivation studies.

Figure 14 shows the measured values and simulated timelines for the cell density (a, c), as well as the substrate and metabolites (b, d). The simulated timelines show

Table 10 Cell growth-dependent parameters used for the simulations of the MC-based hMSC cell growth in the SP100 and SP300

Parameter		Values	Parameter		Values
μ_{max}	1/d	0.64–0.68	*Lac*	mmol/L	0.0
Amn	mmol/L	0.0	q_{Amn}	mmol/cell/d	6–19
Glc	mmol/L	30.5	q_{Glc}	mmol/cell/d	9.8–35
k_{at}	1/d	0.4–1.0	q_{Lac}	mmol/cell/d	20–89
k_{det}	1/d	0.003–0.009	t_l	d	1.5–1.9
K_{Amn}	mmol/L	8–10	X_A	cells/mL	0
K_{Glc}	mmol/L	0.4	X_{Sus}	cells/mL	10,800
K_{Lav}	mmol/L	35–50			

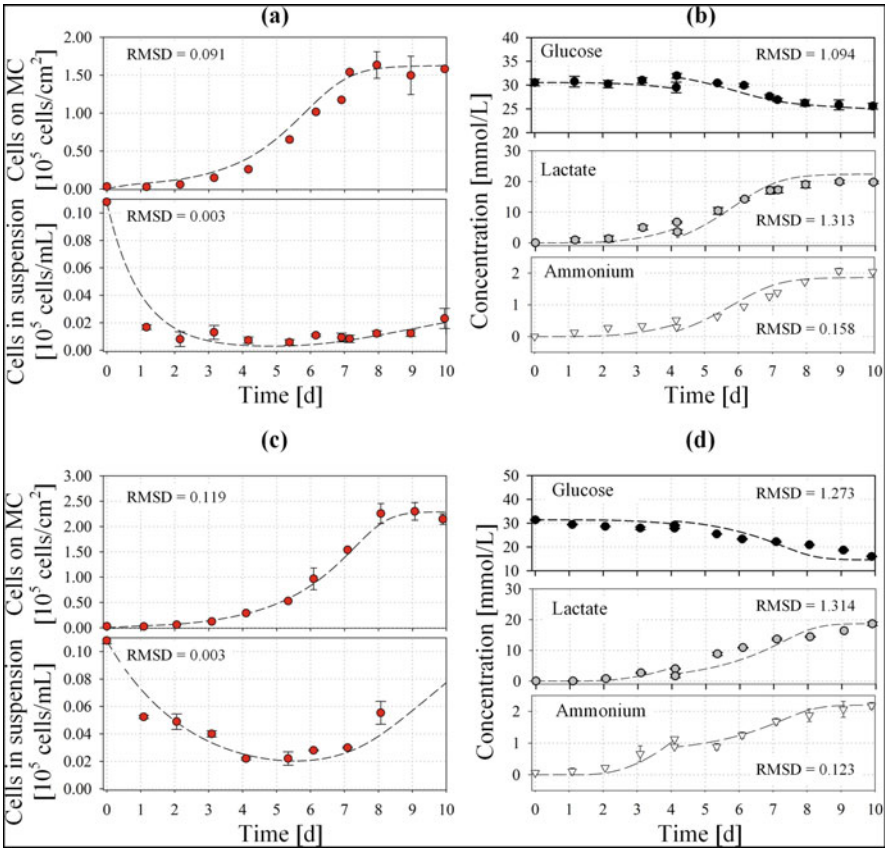


Fig. 14 Comparison of experimental (symbols) and simulated (line) data for cell density (**a**, **c**) and substrate/metabolites (**b**, **d**). The growth simulations were performed for the SP100 (**a**, **b**) and SP300 (**c**, **d**)

pleasing overall correlation with the values measured experimentally and demonstrate the applicability of the unstructured, segregated growth model. By using determined growth parameters from cultivation studies, the cell growth, glucose consumption, lactate production, and ammonium production could be proficiently approximated. The greatest deviations in cell density were in the range of 3–20% for the cells in suspension and 4–24% for the cells on the MCs. The glucose, lactate, and ammonium timelines also correspond to this pattern, even though the specific substrate consumption and metabolite production rates were prone to errors. However, the models provide reliable predictions for the MC-based hMSC growth in the two spinner flask types.

5 Conclusions and Outlook

In this review, the current state of the art of the in vitro expansion of hMSC and the use of numerical tools to support the development of MC-based hMSCs expansions as well as the establishment of "Digital Twins" have been presented. It has been emphasized that different CFD model approaches are described in the scientific literature which can be successfully applied for the characterization of SU bioreactors, especially for the process development of hMSC expansion processes. The CFD case study presented clearly demonstrates that numerical models are valuable tools for the biochemical engineering characterization of small-scale spinner flasks, especially for the determination of parameters that are difficult to determine experimentally. A good correlation was always found between the parameters predicted by the CFD and those measured experimentally. This observation was also in agreement with the literature data. The *Euler-Euler* and *Euler-Lagrange* models gave adequate predictions of the MC distributions within the spinner flask systems and were correlated qualitatively with experimental observations. The *Euler-Lagrange* approach allowed the calculation of particle histories due to its discrete particle formulation, which can be combined with experimental cultivation studies. Thus, *Euler-Lagrange* modelling should be favored in the future in order to derive hydrodynamic stresses over time instead of volume-weighted data. The scientific literature summarized also shows that different model approaches for the simulation of the hMSC growth are available, even though only a few are applicable for the MC-based growth simulation in a stirred bioreactor. The unstructured and segregated growth model presented gives a good description of the MC-based hMSC expansion process in the two spinner flask systems. Thus, MC-based hMSC cell growth can be predicted. However, the further development of descriptive, or even predictive, models for hMSCs will be important in the future for exact scheduling of the preparation of the cell material and the subsequent autologous therapy.

References

1. Grand View Research (2020) Cell therapy market size, share and trends analysis report, 2020–2027
2. Malik NN, Durdy MB (2015) Cell therapy landscape. In: Translational regenerative medicine. Elsevier, pp 87–106
3. Simaria AS, Hassan S, Varadaraju H, Rowley J, Warren K, Vanek P, Farid SS (2014) Allogeneic cell therapy bioprocess economics and optimization: single-use cell expansion technologies. Biotechnol Bioeng 111:69–83. https://doi.org/10.1002/bit.25008
4. Sharma S, Raju R, Sui S, Hu W-S (2011) Stem cell culture engineering – process scale up and beyond. Biotechnol J 6:1317–1329. https://doi.org/10.1002/biot.201000435
5. Ren G, Chen X, Dong F, Li W (2012) Concise review: mesenchymal stem cells and translational medicine: emerging issues. Stem Cells Transl Med 1:51–58
6. Capelli C, Pedrini O, Valgardsdottir R, Da Roit F, Golay J, Introna M (2015) Clinical grade expansion of MSCs. Immunol Lett 168:222–227. https://doi.org/10.1016/j.imlet.2015.06.006
7. Wagner W, Horn P, Castoldi M, Diehlmann A, Bork S, Saffrich R, Benes V, Blake J, Pfister S, Eckstein V, Ho AD (2008) Replicative senescence of mesenchymal stem cells: a continuous and organized process. PLoS One 3:e2213. https://doi.org/10.1371/journal.pone.0002213
8. Lo Surdo J, Bauer SR (2012) Quantitative approaches to detect donor and passage differences in adipogenic potential and clonogenicity in human bone marrow-derived mesenchymal stem cells. Tissue Eng Part C Methods 18:877–889. https://doi.org/10.1089/ten.tec.2011.0736
9. Heathman TRJ, Rafiq QA, Chan AKC, Coopman K, Nienow AW, Kara B, Hewitt CJ (2016) Characterization of human mesenchymal stem cells from multiple donors and the implications for large scale bioprocess development. Biochem Eng J 108:14–23. https://doi.org/10.1016/j.bej.2015.06.018
10. Das R, Roosloot R, van Pel M, Schepers K, Driessen M, Fibbe WE, de Bruijn JD, Roelofs H (2019) Preparing for cell culture scale-out: establishing parity of bioreactor- and flask-expanded mesenchymal stromal cell cultures. J Transl Med 17:241. https://doi.org/10.1186/s12967-019-1989-x
11. Jossen V, Schirmer C, Mostafa Sindi D, Eibl R, Kraume M, Pörtner R, Eibl D (2016) Theoretical and practical issues that are relevant when scaling up hMSC microcarrier production processes. Stem Cells Int 2016:1–15. https://doi.org/10.1155/2016/4760414
12. Jossen V, Eibl R, Kraume M, Eibl D (2018) Growth behavior of human adipose tissue-derived stromal/stem cells at small scale: numerical and experimental investigations. Bioengineering 5:106. https://doi.org/10.3390/bioengineering5040106
13. Hassan S, Simaria AS, Varadaraju H, Gupta S, Warren K, Farid SS (2015) Allogeneic cell therapy bioprocess economics and optimization: downstream processing decisions. Regen Med 10:591–609
14. Lipsitz YY, Milligan WD, Fitzpatrick I, Stalmeijer E, Farid SS, Tan KY, Smith D, Perry R, Carmen J, Chen A, Mooney C, Fink J (2017) A roadmap for cost-of-goods planning to guide economic production of cell therapy products. Cytotherapy 19:1383–1391. https://doi.org/10.1016/j.jcyt.2017.06.009
15. García-Fernández C, López-Fernández A, Borrós S, Lecina M, Vives J (2020) Strategies for large-scale expansion of clinical-grade human multipotent mesenchymal stromal cells. Biochem Eng J 159:107601. https://doi.org/10.1016/j.bej.2020.107601
16. Dolley-Sonneville P, Melkoumian Z, Romeo L. Corning® Stemgro® hMSC Medium. Corning Appl Note 1–8
17. Gottipamula S, Muttigi MS, Chaansa S, Ashwin KM, Priya N, Kolkundkar U, Sundar Raj S, Sen MA, Seetharam RN (2016) Large-scale expansion of pre-isolated bone marrow mesenchymal stromal cells in serum-free conditions. J Tissue Eng Regen Med 10:108–119. https://doi.org/10.1002/term.1713

18. Carter SM, Granchelli J, Stelzer T (2014) Large scale expansion and differentiation of human mesenchymal stromal cells in the Thermo Scientific nunc cell factory system. Thermo Sci Appl Note 1:1–6
19. Jossen V, Muoio F, Panella S, Harder Y, Tallone T, Eibl R (2020) An approach towards a GMP compliant in-vitro expansion of human adipose stem cells for autologous therapies. Bioengineering 7(3):77–100
20. Reichardt A, Polchow B, Shakibaei M, Henrich W, Hetzer R, Lueders C (2013) Large scale expansion of human umbilical cord cells in a rotating bed system bioreactor for cardiovascular tissue engineering applications. Open Biomed Eng J 7:50–61. https://doi.org/10.2174/1874120701307010050
21. Ikebe C, Suzuki K (2014) Mesenchymal stem cells for regenerative therapy: optimization of cell preparation protocols. Biomed Res Int 2014:1–11. https://doi.org/10.1155/2014/951512
22. Scibona E, Morbidelli M (2019) Expansion processes for cell-based therapies. Biotechnol Adv 37:107455. https://doi.org/10.1016/j.biotechadv.2019.107455
23. Discher DE, Mooney DJ, Zandstra PW (2010) Growth factors, matrices, and forces combine. Growth (Lakeland) 324:1673–1677. https://doi.org/10.1126/science.1171643.Growth
24. Steward AJ, Kelly DJ (2015) Mechanical regulation of mesenchymal stem cell differentiation. J Anat 227:717–731. https://doi.org/10.1111/joa.12243
25. Kaiser SC, Eibl D, Eibl R (2015) Single-use bioreactors for animal and human cells. In: Animal cell culture: cell engineering. Springer, Cham, pp 445–499
26. Baraniak PR, McDevitt TC (2012) Scaffold-free culture of mesenchymal stem cell spheroids in suspension preserves multilineage potential. Cell Tissue Res 347:701–711. https://doi.org/10.1007/s00441-011-1215-5
27. Frith JE, Thomson B, Genever PG (2010) Dynamic three-dimensional culture methods enhance mesenchymal stem cell properties and increase therapeutic potential. Tissue Eng Part C Methods 16:735–749. https://doi.org/10.1089/ten.tec.2009.0432
28. Alimperti S, Lei P, Wen Y, Tian J, Campbell AM, Andreadis ST (2014) Serum-free spheroid suspension culture maintains mesenchymal stem cell proliferation and differentiation potential. Biotechnol Prog 30:974–983. https://doi.org/10.1002/btpr.1904
29. Allen LM, Matyas J, Ungrin M, Hart DA, Sen A (2019) Serum-free culture of human mesenchymal stem cell aggregates in suspension bioreactors for tissue engineering applications. Stem Cells Int 2019:1–18. https://doi.org/10.1155/2019/4607461
30. Bhang SH, Cho S-W, La W-G, Lee T-J, Yang HS, Sun A-Y, Baek S-H, Rhie J-W, Kim B-S (2011) Angiogenesis in ischemic tissue produced by spheroid grafting of human adipose-derived stromal cells. Biomaterials 32:2734–2747. https://doi.org/10.1016/j.biomaterials.2010.12.035
31. Layer PG, Robitzki A, Rothermel A, Willbold E (2002) Of layers and spheres: the reaggregate approach in tissue engineering. Trends Neurosci 25:131–134. https://doi.org/10.1016/S0166-2236(00)02036-1
32. Achilli T-M, Meyer J, Morgan JR (2012) Advances in the formation, use and understanding of multi-cellular spheroids. Expert Opin Biol Ther 12:1347–1360. https://doi.org/10.1517/14712598.2012.707181
33. Page H, Flood P, Reynaud EG (2013) Three-dimensional tissue cultures: current trends and beyond. Cell Tissue Res 352:123–131. https://doi.org/10.1007/s00441-012-1441-5
34. Edmondson R, Broglie JJ, Adcock AF, Yang L (2014) Three-dimensional cell culture systems and their applications in drug discovery and cell-based biosensors. Assay Drug Dev Technol 12:207–218. https://doi.org/10.1089/adt.2014.573
35. Caron MMJ, Emans PJ, Coolsen MME, Voss L, Surtel DAM, Cremers A, van Rhijn LW, Welting TJM (2012) Redifferentiation of dedifferentiated human articular chondrocytes: comparison of 2D and 3D cultures. Osteoarthr Cartil 20:1170–1178. https://doi.org/10.1016/j.joca.2012.06.016
36. Bourin P, Bunnell BA, Casteilla L, Dominici M, Katz AJ, March KL, Redl H, Rubin JP, Yoshimura K, Gimble JM (2013) Stromal cells from the adipose tissue-derived stromal

vascular fraction and culture expanded adipose tissue-derived stromal/stem cells: a joint statement of the International Federation for Adipose Therapeutics and Science (IFATS) and the International So. Cytotherapy 15:641–648. https://doi.org/10.1016/j.jcyt.2013.02.006
37. Dominici M, Le Blanc K, Mueller I, Slaper-Cortenbach I, Marini F, Krause D, Deans R, Keating A, Prockop D, Horwitz E (2006) Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. Cytotherapy 8:315–317. https://doi.org/10.1080/14653240600855905
38. Cheng N-C, Chen S-Y, Li J-R, Young T-H (2013) Short-term spheroid formation enhances the regenerative capacity of adipose-derived stem cells by promoting stemness, angiogenesis, and chemotaxis. Stem Cells Transl Med 2:584–594. https://doi.org/10.5966/sctm.2013-0007
39. Bartosh TJ, Ylostalo JH, Mohammadipoor A, Bazhanov N, Coble K, Claypool K, Lee RH, Choi H, Prockop DJ (2010) Aggregation of human mesenchymal stromal cells (MSCs) into 3D spheroids enhances their antiinflammatory properties. Proc Natl Acad Sci 107:13724–13729. https://doi.org/10.1073/pnas.1008117107
40. YlÖstalo JH, Bartosh TJ, Coble K, Prockop DJ (2012) Human mesenchymal stem/stromal cells cultured as spheroids are self-activated to produce prostaglandin E2 that directs stimulated macrophages into an anti-inflammatory phenotype. Stem Cells 30:2283–2296. https://doi.org/10.1002/stem.1191
41. Zimmermann JA, Mcdevitt TC (2014) Pre-conditioning mesenchymal stromal cell spheroids for immunomodulatory paracrine factor secretion. Cytotherapy 16:331–345. https://doi.org/10.1016/j.jcyt.2013.09.004
42. Horn P, Bokermann G, Cholewa D, Bork S, Walenda T, Koch C, Drescher W, Hutschenreuther G, Zenke M, Ho AD, Wagner W (2010) Impact of individual platelet lysates on isolation and growth of human mesenchymal stromal cells. Cytotherapy 12:888–898. https://doi.org/10.3109/14653249.2010.501788
43. Badenes SM, Fernandes TG, Rodrigues CAV, Diogo MM, Cabral JMS (2016) Microcarrier-based platforms for in vitro expansion and differentiation of human pluripotent stem cells in bioreactor culture systems. J Biotechnol 234:71–82. https://doi.org/10.1016/j.jbiotec.2016.07.023
44. Villa-Diaz LG, Ross AM, Lahann J, Krebsbach PH (2013) Concise review: the evolution of human pluripotent stem cell culture: from feeder cells to synthetic coatings. Stem Cells 31:1–7. https://doi.org/10.1002/stem.1260
45. Shearier E, Xing Q, Qian Z, Zhao F (2016) Physiologically low oxygen enhances biomolecule production and stemness of mesenchymal stem cell spheroids. Tissue Eng Part C Methods 22:360–369. https://doi.org/10.1089/ten.tec.2015.0465
46. Wu J, Rostami MR, Cadavid Olaya DP, Tzanakakis ES (2014) Oxygen transport and stem cell aggregation in stirred-suspension bioreactor cultures. PLoS One 9:e102486. https://doi.org/10.1371/journal.pone.0102486
47. Lei Y, Schaffer DV (2013) A fully defined and scalable 3D culture system for human pluripotent stem cell expansion and differentiation. Proc Natl Acad Sci 110:E5039–E5048. https://doi.org/10.1073/pnas.1309408110
48. Sart S, Tsai A-C, Li Y, Ma T (2014) Three-dimensional aggregates of mesenchymal stem cells: cellular mechanisms, biological properties, and applications. Tissue Eng Part B Rev 20:365–380. https://doi.org/10.1089/ten.teb.2013.0537
49. Sucosky P, Osorio DF, Brown JB, Neitzel GP (2004) Fluid mechanics of a spinner-flask bioreactor. Biotechnol Bioeng 85:34–46. https://doi.org/10.1002/bit.10788
50. Kaiser S, Jossen V, Schirmaier C, Eibl D, Brill S, van den Bos C, Eibl R (2013) Fluid flow and cell proliferation of mesenchymal adipose-derived stem cells in small-scale, stirred, single-use bioreactors. Chem Ing Tech 85:95–102. https://doi.org/10.1002/cite.201200180
51. vn Wezel AL (1967) Growth of cell-strains and primary cells on microcarriers in homogeneous culture. Nature 216:64–65

52. Chen AK-L, Reuveny S, Oh SKW (2013) Application of human mesenchymal and pluripotent stem cell microcarrier cultures in cellular therapy: achievements and future direction. Biotechnol Adv 31:1032–1046. https://doi.org/10.1016/j.biotechadv.2013.03.006
53. Jossen V, van den Bos C, Eibl R, Eibl D (2018) Manufacturing human mesenchymal stem cells at clinical scale: process and regulatory challenges. Appl Microbiol Biotechnol 102:3981–3994. https://doi.org/10.1007/s00253-018-8912-x
54. Rafiq QA, Ruck S, Hanga MP, Heathman TRJ, Coopman K, Nienow AW, Williams DJ, Hewitt CJ (2018) Qualitative and quantitative demonstration of bead-to-bead transfer with bone marrow-derived human mesenchymal stem cells on microcarriers: utilising the phenomenon to improve culture performance. Biochem Eng J 135:11–21. https://doi.org/10.1016/j.bej.2017.11.005
55. Leber J, Barekzai J, Blumenstock M, Pospisil B, Salzig D, Czermak P (2017) Microcarrier choice and bead-to-bead transfer for human mesenchymal stem cells in serum-containing and chemically defined media. Process Biochem 59:255–265. https://doi.org/10.1016/j.procbio.2017.03.017
56. Szczypka M, Splan D, Woolls H, Brandwein H (2014) Single-use bioreactors and microcarriers. Bioprocess Int 12:54–64
57. Zhao L-G, Chen S-L, Teng Y-J, An L-P, Wang J, Ma J-L, Xia Y-Y (2014) The MEK5/ERK5 pathway mediates fluid shear stress promoted osteoblast differentiation. Connect Tissue Res 55:96–102. https://doi.org/10.3109/03008207.2013.853755
58. Yim EK, Sheetz MP (2012) Force-dependent cell signaling in stem cell differentiation. Stem Cell Res Ther 3:41. https://doi.org/10.1186/scrt132
59. Frauenschuh S, Reichmann E, Ibold Y, Goetz PM, Sittinger M, Ringe J (2007) A microcarrier-based cultivation system for expansion of primary mesenchymal stem cells. Biotechnol Prog 23:187–193. https://doi.org/10.1021/bp060155w
60. Panchalingam KM, Jung S, Rosenberg L, Behie LA (2015) Bioprocessing strategies for the large-scale production of human mesenchymal stem cells: a review. Stem Cell Res Ther 6:225. https://doi.org/10.1186/s13287-015-0228-5
61. Ferrari C, Balandras F, Guedon E, Olmos E, Chevalot I, Marc A (2012) Limiting cell aggregation during mesenchymal stem cell expansion on microcarriers. Biotechnol Prog 28:780–787. https://doi.org/10.1002/btpr.1527
62. Schirmaier C, Jossen V, Kaiser SC, Jüngerkes F, Brill S, Safavi-Nab A, Siehoff A, van den Bos C, Eibl D, Eibl R (2014) Scale-up of adipose tissue-derived mesenchymal stem cell production in stirred single-use bioreactors under low-serum conditions. Eng Life Sci 14:292–303. https://doi.org/10.1002/elsc.201300134
63. Lawson T, Kehoe DE, Schnitzler AC, Rapiejko PJ, Der KA, Philbrick K, Punreddy S, Rigby S, Smith R, Feng Q, Murrell JR, Rook MS (2017) Process development for expansion of human mesenchymal stromal cells in a 50L single-use stirred tank bioreactor. Biochem Eng J 120:49–62. https://doi.org/10.1016/j.bej.2016.11.020
64. Gruber R, Karreth F, Kandler B, Fuerst G, Rot A, Fischer AB (2004) Platelet-released supernatants increase migration and proliferation, and decrease osteogenic differentiation of bone marrow-derived mesenchymal progenitor cell under in vitro conditions. Platelets 15:29–35
65. Lange C, Cakiroglu F, Spiess AN, Cappallo-Obermann H, Dierlamm J, Zander AR (2007) Accelerated and safe expansion of human mesenchymal stromal cells in animal serum-free medium for transplantation and regenerative medicine. J Cell Physiol 213:18–26
66. Abdelrazik H, Spaggiari GM, Chiossone L, Mretta L (2011) Mesenchymal stem cells expanded in human platelet lysate display a decreased inhibitory capacity on T- and NK-cell proliferation and function. Eur J Immunol 41:3281–3290
67. Heathman TRJJ, Glyn VAM, Picken A, Rafiq QA, Coopman K, Nienow AW, Kara B, Hewitt CJ (2015) Expansion, harvest and cryopreservation of human mesenchymal stem cells in a serum-free microcarrier process. Biotechnol Bioeng 112:1696–1707. https://doi.org/10.1002/bit.25582

68. Carmelo JG, Fernandes-Platzgummer A, Diogo MM, da Silva CL, Cabral JMS (2015) A xeno-free microcarrier-based stirred culture system for the scalable expansion of human mesenchymal stem/stromal cells isolated from bone marrow and adipose tissue. Biotechnol J 10:1235–1247. https://doi.org/10.1002/biot.201400586
69. Rafiq QA, Coopman K, Nienow AW, Hewitt CJ (2016) Systematic microcarrier screening and agitated culture conditions improves human mesenchymal stem cell yield in bioreactors. Biotechnol J 11:473–486. https://doi.org/10.1002/biot.201400862
70. Heathman TRJ, Stolzing A, Fabian C, Rafiq QA, Coopman K, Nienow AW, Kara B, Hewitt CJ (2016) Scalability and process transfer of mesenchymal stromal cell production from monolayer to microcarrier culture using human platelet lysate. Cytotherapy 18:523–535. https://doi.org/10.1016/j.jcyt.2016.01.007
71. Nienow AW, Hewitt CJ, Heathman TRJ, Glyn VAM, Fonte GN, Hanga MP, Coopman K, Rafiq QA (2016) Agitation conditions for the culture and detachment of hMSCs from microcarriers in multiple bioreactor platforms. Biochem Eng J 108:24–29. https://doi.org/10.1016/j.bej.2015.08.003
72. Dufey V, Tacheny A, Art M, Becken U, De Longueville F (2016) Expansion of human bone marrow-derived mesenchymal stem cells in BioBLU 0.3c single-use bioreactors. Appl Note 305:1–8
73. Heathman TRJ, Nienow AW, Rafiq QA, Coopman K, Bo K, Hewitt CJ (2019) Development of a process control strategy for the serum-free microcarrier expansion of human mesenchymal stem cells towards cost-effective and commercially viable manufacturing. Biochem Eng J 141:200–209. https://doi.org/10.1016/j.bej.2018.10.018
74. Jossen V, Kaiser SC, Schirmaier C, Herrmann J, Tappe A, Eibl D, Siehoff A, van d BC, Eibl R (2014) Modification and qualification of a stirred single-use bioreactor for the improved expansion of human mesenchymal stem cells at benchtop scale. Pharm Bioprocess 2:311–322. https://doi.org/10.4155/pbp.14.29
75. Jossen V, Pörtner R, Kaiser SC, Kraume M, Eibl D, Eibl R (2014) Mass production of mesenchymal stem cells – impact of bioreactor design and flow conditions on proliferation and differentiation. In: Eberli D (ed) Cells and biomaterials in regenerative medicine. InTech, Rijeka, pp 119–174
76. Siddiquee K, Sha M (2014) Large-scale production of human mesenchymal stem cells in BioBLU 5c single-use vessels
77. Berry JD, Liovic P, Šutalo ID, Stewart RL, Glattauer V, Meagher L (2016) Characterisation of stresses on microcarriers in a stirred bioreactor. App Math Model 40:6787–6804. https://doi.org/10.1016/j.apm.2016.02.025
78. Paschedag AR (2004) CFD in der Vevfahrenstechnik. Wiley-VCH Verlag GmbH & Co.
79. Ferziger JH, Peric M, Leonard A (1997) Computational methods for fluid dynamics. Phys Today 50:80–84. https://doi.org/10.1063/1.881751
80. Rodriguez S (2019) Applied computational fluid dynamics and turbulence modeling. Springer International Publishing, Cham
81. Delafosse A, Collignon M-L, Marc A, Toye D, Olmos E (2015) Revisiting the determination of hydromechanical stresses encountered by microcarriers in stem cell culture bioreactors. BMC Proc 9:P41. https://doi.org/10.1186/1753-6561-9-S9-P41
82. Liovic P, Šutalo ID, Stewart R, Glattauer V, Meagher L (2012) Fluid flow and stresses on microcarriers in spinner flask bioreactors. Ninth Int Conf CFD Miner Process Ind:1–6
83. Delafosse A, Calvo S, Collignon M-L, Delvigne F, Crine M, Toye D (2015) Euler–Lagrange approach to model heterogeneities in stirred tank bioreactors – comparison to experimental flow characterization and particle tracking. Chem Eng Sci 134:457–466. https://doi.org/10.1016/j.ces.2015.05.045
84. Nienow AW, Rielly CD, Brosnan K, Bargh N, Lee K, Coopman K, Hewitt CJ (2013) The physical characterisation of a microscale parallel bioreactor platform with an industrial CHO cell line expressing an IgG4. Biochem Eng J 76:25–36. https://doi.org/10.1016/j.bej.2013.04.011

85. Collignon M-L, Delafosse A, Calvo S, Martin C, Marc A, Toye D, Olmos E (2016) Large-Eddy simulations of microcarrier exposure to potentially damaging eddies inside mini-bioreactors. Biochem Eng J 108:30–43. https://doi.org/10.1016/j.bej.2015.10.020
86. Kunas KT, Papoutsakis ET (1990) The protective effect of serum against hydrodynamic damage of hybridoma cells in agitated and surface-aerated bioreactors. J Biotechnol 15:57–69. https://doi.org/10.1016/0168-1656(90)90051-C
87. Michaels JD, Petersen JF, Mclntire LV, Papoutsakis ET (1991) Protection mechanisms of freely suspended animal cells (CRL 8018) from fluid-mechanical injury. Viscometric and bioreactor studies using serum, pluronic F68 and polyethylene glycol. Biotechnol Bioeng 38:169–180. https://doi.org/10.1002/bit.260380209
88. Chisti Y (2000) Animal-cell damage in sparged bioreactors. Trends Biotechnol 18:420–432. https://doi.org/10.1016/S0167-7799(00)01474-8
89. Jossen V (2020) Bioengineering aspects of microcarrier-based hMSC expansions in different single-use bioreactors. Technical University of Berlin, Berlin
90. Stoots CM, Calabrese RV (1995) Mean velocity field to a rushton turbine blade. Am Inst Chem Eng J 41:1–11
91. Wollny S (2010) Experimentelle und numerische Untersuchungen zur Partikelbeanspruchung in gerührten (Bio-)Reaktoren. Technical University of Berlin
92. Venkat RV, Stock LR, Chalmers JJ (2000) Study of hydrodynamics in microcarrier culture spinner vessels: a particle tracking velocimetry approach. Biotechnol Bioeng 49:456–466. https://doi.org/10.1002/(SICI)1097-0290(19960220)49:4<456::AID-BIT13>3.0.CO;2-8
93. Ismadi M-Z, Hourigan K, Fouras A (2014) Experimental characterisation of fluid mechanics in a spinner flask bioreactor. Processes 2:753–772. https://doi.org/10.3390/pr2040753
94. Zhang H, Lamping SR, Pickering SCR, Lye GJ, Shamlou PA (2008) Engineering characteristics of a single well from 24-well and 96-well microtire plates. Biochem Eng J 40:138–149
95. Godara P, McFarland CD, Nordon RE (2008) Design of bioreactors for mesenchymal stem cell tissue engineering. J Chem Technol Biotechnol 83:408–420. https://doi.org/10.1002/jctb.1918
96. Rafiq QA, Brosnan KM, Coopman K, Nienow AW, Hewitt CJ (2013) Culture of human mesenchymal stem cells on microcarriers in a 5 l stirred-tank bioreactor. Biotechnol Lett 35:1233–1245. https://doi.org/10.1007/s10529-013-1211-9
97. Ibrahim S, Nienow AW (2004) Suspension of microcarriers for cell culture with axial flow impellers. Chem Eng Res Des 82:1082–1088. https://doi.org/10.1205/cerd.82.9.1082.44161
98. Hewitt CJ, Lee K, Nienow AW, Thomas RJ, Smith M, Thomas CR (2011) Expansion of human mesenchymal stem cells on microcarriers. Biotechnol Lett 33:2325–2335. https://doi.org/10.1007/s10529-011-0695-4
99. Yourek G, McCormick SM, Mao JJ, Reilly GC (2010) Shear stress induces osteogenic differentiation of human mesenchymal stem cells. Regen Med 5:713–724. https://doi.org/10.2217/rme.10.60
100. Yourek G, Hussain MA, Mao JJ (2007) Cytoskeletal changes of mesenchymal stem cells during differentiation. ASAIO J 53:219–228. https://doi.org/10.1097/MAT.0b013e31802deb2d
101. Yeatts AB, Choquette DT, Fisher JP (2013) Bioreactors to influence stem cell fate: augmentation of mesenchymal stem cell signaling pathways via dynamic culture systems. Biochim Biophys Acta Gen Subj 1830:2470–2480. https://doi.org/10.1016/j.bbagen.2012.06.007
102. Yeatts AB, Fisher JP (2011) Bone tissue engineering bioreactors: dynamic culture and the influence of shear stress. Bone 48:171–181. https://doi.org/10.1016/j.bone.2010.09.138
103. Weyand B, Reimers K, Vogt PM (2011) Influences of extracellular matrix properties and flow shear stresses on stem cell shape in a three-dimensional dynamic environment. IFMBE Proc 30:47–50
104. Weyand B, Kasper C, Israelowitz M, Gille C, von Schroeder HP, Reimers K, Vogt PM (2012) A differential pressure laminar flow reactor supports osteogenic differentiation and

extracellular matrix formation from adipose mesenchymal stem cells in a macroporous ceramic scaffold. Biores Open Access 1:145–157
105. Weyand B, Israelowitz M, von Schroeder HP, Vogt PM (2009) Fluid dynamics in bioreactor design: considerations for the theoretical and practical approach. Adv Biochem Eng Biotechnol 112:251–268
106. Croughan MS, Hamel J-F, Wang DIC (2006) Hydrodynamic effects on animal cells grown in microcarrier cultures. Biotechnol Bioeng 95:295–305. https://doi.org/10.1002/bit.21158
107. Heathman TRJ, Stolzing A, Fabian C, Rafiq QA, Coopman K, Nienow AW, Kara B, Hewitt CJ (2015) Serum-free process development: improving the yield and consistency of human mesenchymal stromal cell production. Cytotherapy 17:1524–1535. https://doi.org/10.1016/j.jcyt.2015.08.002
108. Cierpka K, Elseberg CL, Niss K, Kassem M, Salzig D, Czermak P (2013) hMSC production in disposable bioreactors with regards to GMP and PAT. Chem Ing Tech 85:67–75. https://doi.org/10.1002/cite.201200151
109. Rafiq QA, Hanga MP, Heathman TRJ, Coopman K, Nienow AW, Williams DJ, Hewitt CJ (2017) Process development of human multipotent stromal cell microcarrier culture using an automated high-throughput microbioreactor. Biotechnol Bioeng 114:2253–2266. https://doi.org/10.1002/bit.26359
110. Higuera G, Schop D, Janssen F, van Dijkhuizen-Radersma R, van Boxtel T, van Blitterswijk CA (2009) Quantifying in vitro growth and metabolism kinetics of human mesenchymal stem cells using a mathematical model. Tissue Eng Part A 15:2653–2663. https://doi.org/10.1089/ten.tea.2008.0328
111. dos Santos F, Andrade PZ, Boura JS, Abecasis MM, da Silva CL, Cabral JMS (2009) Ex vivo expansion of human mesenchymal stem cells: a more effective cell proliferation kinetics and metabolism under hypoxia. J Cell Physiol 223:n/a–n/a. https://doi.org/10.1002/jcp.21987
112. Bartolini E, Manoli H, Costamagna E, Jeyaseelan HA, Hamad M, Irhimeh MR, Khademhosseini A, Abbas A (2015) Population balance modelling of stem cell culture in 3D suspension bioreactors. Chem Eng Res Des 101:125–134. https://doi.org/10.1016/j.cherd.2015.07.014
113. Mancuso L, Ilaria Liuzzo M, Fadda S, Cincotti A, Pisu M, Concas A, Cao G (2010) Experimental analysis and modeling of bone marrow mesenchymal stem cells proliferation. Chem Eng Sci 65:562–568. https://doi.org/10.1016/j.ces.2009.06.034
114. Bailón-Plaza A, van der Meulen MCH (2001) A mathematical framework to study the effects of growth factor influences on fracture healing. J Theor Biol 212:191–209. https://doi.org/10.1006/jtbi.2001.2372
115. Geris L, Peiffer V, Demol J, Oosterwyck H Van (2006) Modelling of in vitro mesenchymal stem cell cultivation, chondrogenesis and osteogenesis. J Biomech 41:466–466
116. Schellenberg A, Stiehl T, Horn P, Joussen S, Pallua N, Ho AD, Wagner W (2012) Population dynamics of mesenchymal stromal cells during culture expansion. Cytotherapy 14:401–411. https://doi.org/10.3109/14653249.2011.640669
117. Cholewa D, Stiehl T, Schellenberg A, Bokermann G, Joussen S, Koch C, Walenda T, Pallua N, Marciniak-Czochra A, Suschek CV, Wagner W (2011) Expansion of adipose mesenchymal stromal cells is affected by human platelet lysate and plating density. Cell Transplant 20:1409–1422. https://doi.org/10.3727/096368910X557218
118. Hoffmann M, Kuska J-P, Zscharnack M, Loeffler M, Galle J (2011) Spatial organization of mesenchymal stem cells in vitro – results from a new individual cell-based model with podia. PLoS One 6:e21960. https://doi.org/10.1371/journal.pone.0021960

Adv Biochem Eng Biotechnol (2021) 177: 229–254
https://doi.org/10.1007/10_2020_133

Published online: 26 September 2020

Euler-Lagrangian Simulations: A Proper Tool for Predicting Cellular Performance in Industrial Scale Bioreactors

Christopher Sarkizi Shams Hajian, Julia Zieringer, and Ralf Takors

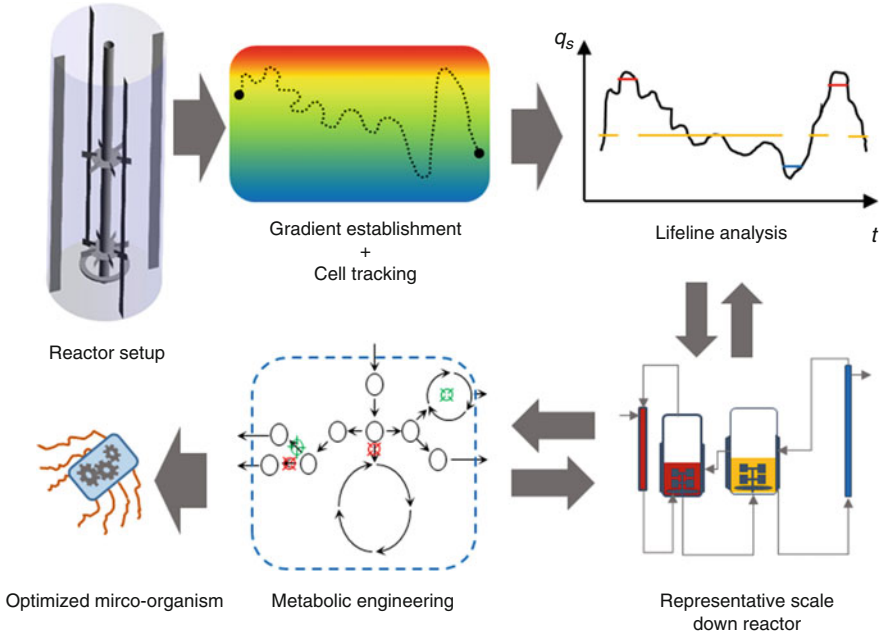
Contents

Abstract Eulerian-Lagrangian approach to investigate cellular responses in a bioreactor has become the center of attention in recent years. It was introduced to biotechnological processes about two decades ago, but within the last few years, it proved itself as a powerful tool to address scale-up and -down topics of bioprocesses. It can capture the history of a cell and reveal invaluable information for, not only, bioprocess control and design but also strain engineering. This way it will be possible to shed light on the actual environment that cell experiences throughout its lifespan. Lifelines of a microorganism in a bioreactor can serve as the missing link that encompasses the biological timescales and the physical timescales. For this purpose digitalization of bioreactors provides us with new insights that are not

C. S. S. Hajian, J. Zieringer, and R. Takors (✉)
Institute of Biochemical Engineering, University of Stuttgart, Stuttgart, Germany
e-mail: takors@uni-stuttgart.de

achievable in industrial reactors easily if at all, namely, substrate and product gradients; high-shear regions are among the most interesting factors that can be reproduced adequately with help of a digital twin. In this chapter basic principles of this method will be introduced, and later on some practical aspects of particle tracking technique will be illustrated. In the final section, some of the advantages and challenges associated with this method will be discussed.

Graphical Abstract



Keywords Cell life-lines, Digital twins, Large-scale bioprocesses, Scale-down, Scale-up

1 Introduction

The physiological state of microorganisms and its impact on growth and product formation are the result of complex interactions between the cellular environment and the cells. Large-scale studies have shown that homogeneous culture conditions are difficult to establish. This is expressed in the common correlation for stirred tank reactors indicating that mixing time τ_{mix} is proportional to $\left(\frac{P}{V}\right)^{-1/3}$ (with P as power input and V as volume) pinpointing to increasing mixing times with reduced power-to-volume inputs. The latter occurs typically in large-scale bioreactors because of limiting power supply, engines, gearings, etc. Nevertheless process engineering and

bioreactor design may aim to create the least heterogeneous impact possible [1]. Independent of the bioreactor type at hand, heterogeneities happen in large scale, and as a result product loss seems inevitable while scaling up. Heterogeneities expose cell to various stresses which will in turn translate to product loss. Trail experiments in actual industrial bioreactors are not feasible at all, and that is one of the main motivations of a "digital twin" of the bioreactor.

A digital twin allows us to investigate and locate the physical and biological bottleneck. It enables the user to predict various scenarios in parallel saving significant resources. Once the digital twin of the bioreactor is validated, it can be translated to a scale-down bioreactor that represents the imperfections of the larger scale. This setup can be investigated to reveal the most significant biological behaviors of the cell and ultimately providing the digital twin for the cell.

In theory once digital twins for the cell and for the bioreactor are available, versatile optimization of bioprocesses would be carried out in only a fraction of the time and resource. This offers the operator the choice between improving the bioreactor, microorganism, and both.

2 Embedding Cells in Microenvironmental Heterogeneities of Bioreactors

To predict cellular responses in large-scale bioreactors, two crucial prerequisites are needed: (1) models simulating spatially resolved substrate availabilities, flows, and mass transfer and (2) models translating microenvironmental heterogeneities into proper cellular responses. Research of the last years enabled substantial improvements in both fields, computational fluid dynamics and thorough experimental studies, thereby providing the ground for large-scale prediction of microbial performance a priori.

Large-scale bioreactor conditions need to be calculated, aiming at a spatial resolution of mass, momentum, and energy balances via numerical simulations. In particular, Navier-Stokes (NSE) and continuity equations representing the conservation of momentum and mass should be solved. Basically, NSEs describe the motion of viscous fluid flows with the fluids considered as a continuum rather than colliding particles. Under the typical mixing conditions given, the occurrence of turbulent zones is likely. They are integrated via additional transport equations typically demanding for two additional equations depending on the models applied. Turbulence is provoked by eddies which affect molecular diffusion, heat transfer, and the mixing behavior. For flows involving heat transfer or compressibility, an additional equation for energy conservation should be solved. Furthermore, the balancing of individual species (particles or cells) that undergo mixing and reactive changes requires the implementation of proper conservation terms. Depending on the available computing power, simulations can range from direct numerical simulation (DNS) via large eddy simulation (LES) to Reynolds-averaged Navier-Stokes

(RANS) approaches. DNS yields to solve each individual turbulent structure, whereas LES only directly solves the large energy-containing scales, while the effects of the more universal small scales are modeled. RANS comprises time-averaged flow equations which allow to simulate small- and large-scale eddies with a minimum but still challenging computational efforts (for more details, see [2]). Consequently, RANS simulations are often favored. They only require 1/100 to 1/10 computational efforts compared to LES [3] or DNS. Although RANS models require several modeling assumptions and approximations, their predictive power is sufficient for providing insight in reactor-scale substrate concentration gradients [4].

Many research projects have shown that cells react in a multiresponse, multilayer fashion comprising the on- and offset of transcriptional regulation programs, as well as proteomic and metabolic changes [5–11]. The latter are subject of state-of-the-art approaches mirroring the instantaneous metabolic response on extracellular heterogeneities [12–16]. However, the consideration of transcriptional and translational effects introduces different timescales of cellular response. Consequently, initiation and execution may be spatially disconnected which differs fundamentally from the instantaneous metabolic responses studied so far [17, 18]. Coupling cellular kinetics with fluid dynamics consolidates the knowledge about microbial cell kinetics with fluid dynamics and mass transfer in industrial scale bioreactors yielding to optimize the design of both [16].

To investigate the consequences of environmental heterogeneities, proper modeling frameworks should link local variations with cellular and subcellular kinetics. Mainly two different methods exist in computational fluid dynamic (CFD) to display the adaptation of the cellular behavior to the environment: (1) population balance models (PBM) [19–21] and (2) the Euler-Lagrangian method (EL) [12, 13, 15, 22–25]. Both methods rely on the grid-based Euler approach to simulate the continuum background, meaning the fluid surroundings of the particles. The grid-based method relies on spatial volume discretization. Adjacent volumes are connected via transport equations and are considered as homogeneous. Consequently, the following rule of thumb holds true: the smaller the volumes or the finer the grid, the more realistic is the output. An extensive study of mesh dependency on biological output has been conducted by Kuschel and Takors [26].

In the PBM approach, microorganisms are considered as part of the continuum with no erratic changes allowed [19, 27, 28]. Particles are grouped in classes, and a predefined distribution range of particles is implemented via distribution density functions. These equations are useful to determine relevant macroscopic properties such as the interfacial area or the biomass-specific growth rate [12, 14, 29]. Hence, PBMs represent a powerful modeling framework for the description of fundamental properties that are characterized by distributions in a coarse timescale. However, cellular adaptations may happen on different timescales than macroscopic fluctuations and may show much more interactions than implemented in common PBMs.

To overcome those limitations, the Lagrangian method may be applied. Individual properties are assigned to each moving particle (e.g., biological cell). Noteworthy, these traits cause interactions with the environment receiving equal environmental feedback. In essence, Euler-Lagrange approach (EL) tracks a given

number of particles while fluctuating in the bioreactor, thereby recording interactions with the environment. The dispersed phase can exchange momentum, mass, and energy with the fluid phase. The approach renders considerably simpler when particle-particle interactions may be neglected. This requires a fully diluted (volume fraction $<10\%$) dispersed second phase. Considering the case of biological cells as moving "particles," further simplifications are often assumed: As the observation window of fluctuating cells is smaller than time constants of physical changes inside the grids, cellular impacts on physical states inside the grid are often neglected.

Trajectories of an individual particle are predicted by integrating particle forces which is coded in a Lagrangian reference frame. In essence, particle inertia is balanced with forces acting on the particle. Typically, the Stokes number for a microorganism (diameter: 5×10^{-6} m) is <0.01 giving rise to the fair assumption that particles move with the flow possessing a negligible mass [12, 14]. Hence, massless particles can be treated as ideal flow followers and immediately adapt to the local flow velocity, and no force balance has to be solved, reducing the computation time considerably [30]. However, if bulk velocity is the only motional source of particles, they tend to get stuck in poorly or steadily mixed zones. To be precise, in order to prevent trapping close to reactor surfaces or inside eddies, particle-turbulence is modeled via the discrete random walk (DRW). Mimicking characteristic circulation times in small eddies, random velocity is activated to enable particle escape. The approach leads to temporary constant functions mirroring fluctuating velocity impacts [30, 31].

A major challenge of the EL method is the significant computational burden. As discussed in [32, 33], the computation time depends on the required time resolution and the number of tracked particles. First, the broad range of timescales for metabolic reaction and cellular adaptation [6, 17, 18] requires simulation of flow with fine temporal resolution. For instance, time steps of milliseconds are needed to track particles properly especially in highly agitated flows. Second, a large number of particles may be required to fulfill the so-called ergodicity constraint, a prerequisite for a sound biological readout. Too low numbers of tracked particles lead to artificial spatial variations finally causing nonrealistic interpretations.

PBM and EL frameworks are usually applied for monophasic conditions represented by the Euler phase. The dispersed phase typically comprises either bubbles or particles, but not both in combination. However, tracking individual cells in bubbled (aerated) bioreactors should be an attractive goal for future application to analyze aerobic cultivations [11]. This may require a combination of both methods outlined above, leading to an Euler-Euler-Lagrange formulation (EEL). In detail, the fluid surrounding the bubbles and particles is calculated via the Euler approach, while the interfacial area of the bubbles (for mass transfer) is derived with the population balance method. Additionally, the history of cells may be recorded via Lagrangian particle tracking.

2.1 The Core Idea of Lifeline Analysis

After different methods have been addressed in the previous section, the focus is now on implementing the EL framework in CFD. The basis for each simulation is the geometry and the spatial discretization of its volume. To obtain high-quality discretization, a so-called mesh, the geometry needs to be simplified by removing edges, surface filets, and nonessential components. Accordingly, agitators and baffles may be replaced by zero-thickness walls [34]. Commercial codes often make use of orthogonal quality, aspect ratio, and skewness of individual grid cells to qualify the mesh.

To simulate the continuum environment of a bioreactor, the Eulerian approach is used together with a case-specific model. Most commonly, the RANS standard k-ε model is chosen for stirred tank reactors [12, 14, 15, 35, 36]. As known, standard RANS models work well in flow scenarios with lowly frequent changing mean flows compared to high turbulence frequencies. However, the turbulent dissipation rate (ε) is underestimated near the impeller tip but shows decent agreement in the bulk [37]. Consequently, new methods of turbulence modeling combining advantages of different models have been proposed, recently [38, 39]. In essence, they apply zonal methods such as the RANS/LES approach. RANS and a subgrid-scale model (LES) are used in different domains separated by sharp or dynamic interfaces. In general, such hybrid models are useful to increase the accuracy where necessary but keeping computational demands limited.

Convergence of the solution can be assumed if all residuals are below 10^{-5} and oscillation in the mean velocity magnitude and bubble diameter (Parameter of interest depend on the individual simulation) is below 1% [12, 14]. When steady state is reached zero-thickness walls can be converted from wall to interface/interior boundary conditions to prevent particle trapping when the frozen flow field assumption is made. Although the solution is converged, the results might not be accurate. To improve the accuracy of the simulation, the results need to be mesh-independent. That means the results (e.g., ε, power number or velocity magnitude for stirred tank reactors) do not vary significantly (below 5%) when the mesh density is increased. Now, the mesh-independent mixing time may be used to validate the resulting flow patterns, the simulated impeller power numbers, k, and ε profiles in the discharge stream of the impeller. Noteworthy, simulated mixing times of non-aerated scenarios are expected to be shorter than true mixing times of aerated processes.

As a result, a steady-state converged flow field is calculated that is not updated during the particle tracking phase. Basically, the procedure mirrors the simplifying assumption of massless cells neither interacting with each other nor with the continuum surrounding. Accordingly, the history of cellular experiences recorded in particle lifelines only monitors the interaction with fixed hydrodynamics and concentration profiles. The latter may be established using black-box cellular uptake kinetics [12, 14, 15]. Using nonstructured Monod-type kinetics, the biomass-specific substrate uptake rate instantaneously adapts to the local concentrations. Hence, reaction kinetics may even be embedded in the liquid phase to save computation

time. In essence, this simplification mirrors the assumption of homogeneously dispersed bacteria [12, 14]. Other more complex approaches using a structured kinetic model coupled to the gradient in the Euler phase to directly determine the intracellular state of the microbial cell have been applied but are limited due to their computational demand [30, 36]. Monod-type kinetics are mainly valid for long-term balanced growth which might hold true for a lab-scale chemostat to characterize a strain but falls short when offering practical solution for industrial processes like a substrate limited fed-batch [40–42]. In the case of overflow metabolism for baker's yeast, several structured models were employed in various scenarios some of which include glycolytic oscillations in a population of yeast cells [36], continuous fermentation of *Saccharomyces cerevisiae* [41], fed-batch fermentation with emphasis on ethanol concentration of oxygen uptake rate, and questioning pyruvate dehydrogenase's role as a bottleneck enzyme [43], a cybernetic model for batch fermentation [44]. Availability of such information encourages researchers to employ more sophisticated kinetic models in their CFD simulations [32, 45].

To qualify whether or not statistically sufficient particles (cells) are tracked, the ergodic theorem is typically applied. The criterion follows the fundamental idea that a stochastic process proceeding in time converges to the same average as the average of the entire system's space. In other words, averages of state variables recorded in lifelines should be the same irrespective whether one single particle is tracked for endless time or a proper number is monitored for shorter intervals. In practice, typically 10^5 particles are added (well distributed in space) and mixed, at least for the duration of the mixing time to ensure a homogeneous distribution. Individual experiences of each particle are recorded for further analysis.

Depending on the goal and timescale of the process, the readout frequency must be appropriate to provide reasonable resolution (usually in the range of 10–30 ms). As outlined above, particle slip with the convective fluid may be neglected which basically reflects the very low Archimedes numbers of microbial cells [12, 14]. By analogy, momentum transfer between the particles and the fluid phase is not considered either, again reflecting the small masses of the cells (quasi-single phase, [46]). Accordingly, so-called one-way coupled EL approaches are today's standard. On contrast, two-way coupled simulations considering the particle-continuum interactions require very high computational demands that are hardly applicable yet [47].

Even with many assumptions made, the system of equations resulting from discretization is tremendous due to the facts that (a) it is a three-dimensional problem and (b) a sufficiently large number of particles and time steps are required to achieve a realistic and statistically sound description of the population. To keep the computational cost-feasible assumptions on operation type of the process, kinetics are taken into account. To reduce the computational burden for simulation, multiphase simulation is done in Eulerian frame [12, 14, 15, 36], or if it is known that oxygen is found in excess quantities all over reactor and its volume fraction does not exceed 10%, simpler approaches can be employed [13, 22–24]. Currently, the common approach is to apply one-way coupling meaning that particles are affect by the flow field and not the other way around [15, 25, 36]. Haringa et al. [13, 22–24] considered two-way

coupling only for substrate concentration. As for the particle number, ergodic theorem is explained later on in the chapter. All these assumptions should be based on prior knowledge and by having the end in mind as too little detail will result in inaccuracy and too much detail will result in unfeasible resource utilization.

2.2 How to Get Biologically Sound Readouts?

After tracking individual particles and their properties, the analysis may be performed in a different environment, such as MATLAB. For clustering particle properties, proper regimes must be defined first. For instance, critical substrate concentration may be set as thresholds distinguishing different metabolic states. Or critical ε values and critical ε frequencies may be defined to qualify shear stress for susceptible cultures. Another point of view is to integrate a more complicated kinetic model based on multiple substrates and by-product lifelines taking exposure statistics into account which is an interesting view when investigating overflow metabolism [26] and many more, etc. In combination with regime constraints, residence time distributions of cells resting within the defined regime borders should be studied, too [12, 14, 15].

Often, turbulent movements reveal very rapid regime changes for those particles possessing properties contiguous to the rigid regime borders. The resulting artificial regime changes bias residence time distributions and should be curated, accordingly. Haringa et al. [12, 14] distinguished between three types of rapid variations: rapid successive regime crossings, integral scale variations, and subgrid variations by small eddies. Based on the work of Linkès et al. [48], eddy micro-mixing is not limited in nonviscous fluids, which translates to assimilation timescale being orders of magnitude larger than the timescale of the smallest turbulent structures which is referred to as Kolmogorov timescale $\left(\frac{\nu}{\varepsilon}\right)^{0.5}$ and therefore concluding that micro-mixing is not a limiting step considering substrate assimilation (no effect on gradient formation). The other two effects occur at the Lagrangian timescale when applying the DRW model and might have an effect on substrate assimilation and gradient formation. For a tractable analysis, the turbulent variations need to be smoothed using appropriate filters, with a filter time step corresponding to the Lagrangian timescale.

Alternately, to soften the hard boundary conditions (which do not exist in reality) a second "boundary" (fuzzy) filter can be applied (e.g., filter amplitude of ± 0.01 in concentration changes). Besides, investigating the influence of highly fluctuating substrate availability on the cellular performance may be an interesting research topic for future studies.

In general, sound knowhow about cellular metabolism and regulation is mandatory to qualify the readout of lifelines. From a fluid mechanics point of view in reality, a fluctuation in a concentration can be as small as the Kolmogorov scales, yet it is only limited to the time resolution of the digital twin of the bioreactor. The same

holds true for the microorganism even with the extracellular fluctuations calculated; the cellular response can only be representative if and only if it is based on real-life behavior of the cell. The fact that the macro-twin (digital bioreactor) and micro-twin (digital microorganism) go hand in hand has another advantage other than being representative. The timescales indicate how fine the resolution of the computations should be and hence allowing to a have an idea of what resources would be needed beforehand. The main traits of interest in the biological side are associated with the sensitivity of the organism to different occurrences and how fast and on what level the response takes place. Another aspect to focus on is to evaluate the response initiation and termination. Once this is set, the cellular machinery involved can be included in the micro-twin with hopes of shedding light on the effect of complex biological features on an industrial bioprocess. Accordingly, proper data are crucially needed to develop a thorough mechanistic understanding [6, 11, 13, 49].

3 Lifeline Analysis in Practice

In this section readers will be introduced to an exemplified bioprocess for creating a holistic idea of this particular EL concept in a practical manner. This will serve as a foundation to highlight strengths and bottlenecks presented by the technique. It is worth noting that the simulation environment (here: ANSYS Fluent 2019 R1) might have some practical superiority to other software packages, but the basics remain the same and are not software specific.

Figure 1 aims to illustrate how interconnected elements of a multiphase simulation can be thereby disclosing the complexity of the network. On the one hand, one

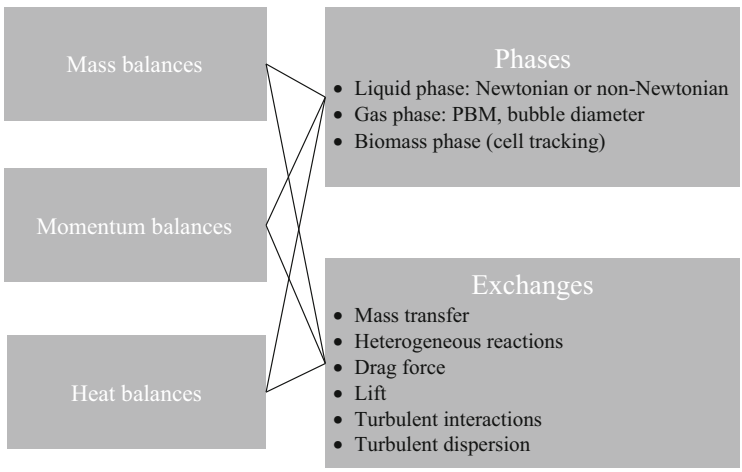


Fig. 1 Various sets of interacting equations which are of interest in a typical CFD simulation of a multiphase bioreactor

may yield to describe finest details which not only cause highest computational effort but also highest likeliness of numerical instabilities and inaccuracies. On the other hand, less complication equals less computational resources required.

Real applications always have to decide between those extremes finding the proper compromise for the given problem. Examples of such simulations are abundant in the literature and will not be further discussed here [2, 13, 15, 25, 46, 50, 51].

To realize the importance of this approach, we find it beneficial for the readers to get familiar with the procedures of the method. Gradients are one of the most significant aspects of large-scale fermentations. In first part the Eulerian gradients are resolved and then the fate of the cells within these gradients is investigated.

3.1 Eulerian Simulation Setup

CFD simulations are carried out using ANSYS Fluent 2019 R1, and post processing is conducted. The chosen bioreactor has 54 m^3 equipped with two different Rushton turbines (Fig. 2). The impeller on the bottom has eight blades, whereas the one on top possesses six. For a more detailed schematic, refer to [13, 22–24].

k-ε models are vastly used in the industry to simulate stirred tank reactors because of their low burden on computational power needed and sufficient accuracy that they provide. In this example, Eulerian model is used for multiphase simulation; for the sake of simplicity, the broth is assumed to have characteristics of water, and for the gas phase, air is chosen. Based on Haringa et al. [13, 22–24], a single bubble diameter of 7 mm is set for the gas phase.

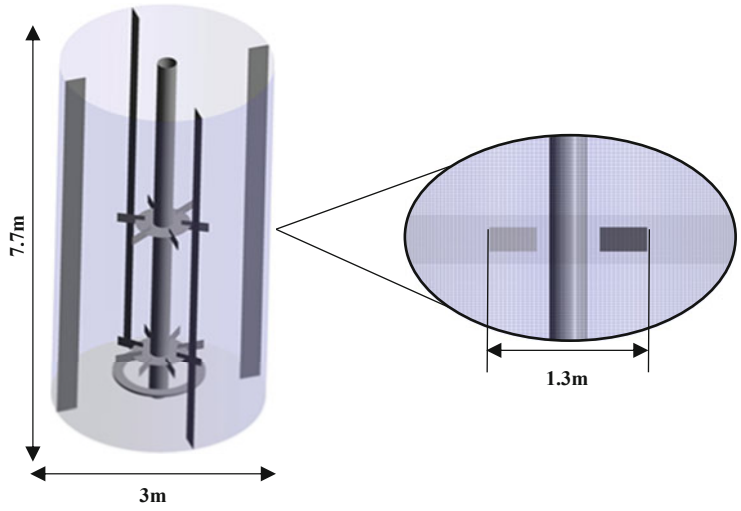


Fig. 2 Geometry of the investigated aerated baffled bioreactor prepared grid of 1,500,000 hexahesdrons for simulation

To take drag force into account, the so-called Grace model [52] is employed which tends to give a mathematically stable solution by offering physically acceptable gas flow regimes.

For grid independence, three different meshed geometries with 500,000, 1,000,000, and 1,500,000 were compared based on the differences in turbulent parameters k and ε (data not shown here). The procedure includes parameter evaluation nearby the impeller, a location of outmost importance in stirred bioreactors. If the difference of aforementioned values is less than an acceptable threshold (in our case 5%), the simulation is considered acceptably grid independent.

Because of the complex and transient flow fields reaching the usual thresholds of maximum residuals, $<10^{-6}$ is not possible without refining computational time steps for a few orders of magnitude. The latter increases computational times almost proportionally. Usually, as done here, pseudo-steady states are defined when solutions converge below the given threshold and residuals of continuity fluctuate within 10^{-3}–10^{-4}. To keep the setup simple, a moving reference frame (MRF) is employed to set the agitation rate at 150 rpm.

3.2 Eulerian Simulation Outputs

Considering three probes (bottom, middle, top), for example, mixing times are estimated about 40s as shown in Fig. 3. The oxygen transfer rate (1) is estimated using the model of Lamont & Scott [53] to calculate the mass transfer coefficient (2):

$$\mathrm{OTR} = k_L \times a \left(C^*_{O_2} - C_{O_2} \right) \tag{1}$$

where

$$k_l = 0.4 \times \sqrt{D} \times \left(\frac{\varepsilon}{\nu} \right)^{0.25} \tag{2}$$

ε is calculated from the CFD simulation, and the kinematic viscosity ν is that of water which is around 10^{-6} m^2/s at 25°C. D is the diffusion coefficient of oxygen in water, 2.1×10^{-9} m^2/s. Applying Henry's law, all the variables needed to calculate the OTR are accessible.

A variation of Monod kinetics [54] is assumed (3) to describe oxygen and glucose dependency of growth (Table 1). In essence, growth is limited by the lowest availability of each component. This approach performs better compared to models which include the effect of multiple substrates by multiplication. This measure is taken to avoid under prediction of growth and uptake rates where both substrates are limiting.

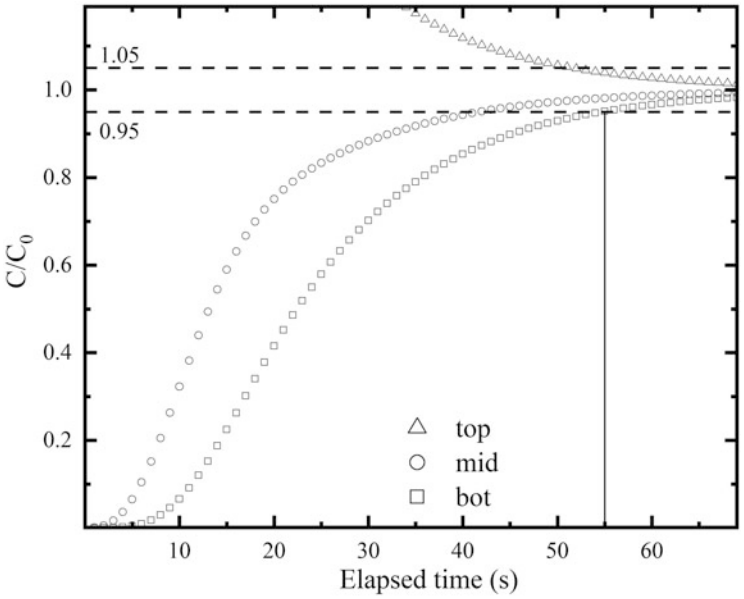


Fig. 3 Mixing time estimation using three different probe locations (bottom) 0.2 m, (middle) 3.8 m, and (top) 7 m from the bottom of the tank

Table 1 Kinetic parameters used to model substrate and oxygen uptake rates

Kinetic parameter	Value	Ref.
K_s	22.2 μmol · l^{-1}	[55]
K_{O_2}	1.8 μmol · l^{-1}	[56]
$\mu_{\max}$	0.5 h^{-1}	[57]
$q_{O_2,\max}$	0.5 g · $g_x^{-1}h^{-1}$	[58]
Y_{xs}	0.5 g_x · g^{-1}	[57]
C_x	55 g_x · l^{-1}	[59]

$$\mu = \mu_{\max} \times \mathrm{Min}\left(\frac{C_{\mathrm{Glucose}}}{K_{\mathrm{Glucose}} + C_{\mathrm{Glucose}}}, \frac{C_{O_2}}{K_{O_2} + C_{O_2}}\right) \tag{3}$$

Glucose feed and gas flow rates are 0.11 mol/s and 0.21 kg/s, respectively. These values are handpicked to put emphasis on the underlying gradients affecting the specific regimes. Here, we characterize the microbial kinetics for a hypothetical *E. coli* strain with respect to available glucose and oxygen concentrations and assign regimes (Fig. 4).

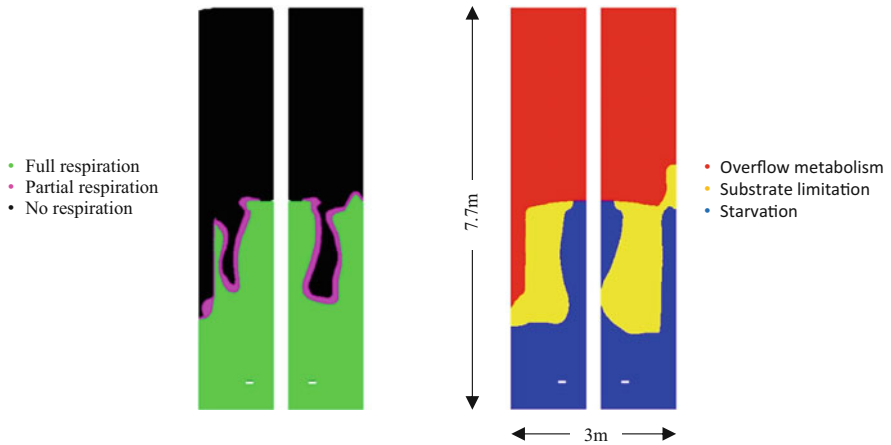


Fig. 4 Eulerian simulation outputs: oxygen concentration regimes indicating respiration in *green*, partial respiration in *magenta*, anaerobic conditions in *black* (*left*), glucose concentration regimes where overflow metabolism takes place are indicated with *red*, limitation zone with *yellow*, and starvation compartment with *blue* (*right*)

3.3 Lagrangian Setup

To introduce the Lagrangian phase, we should indicate the interactions between the particles (here cells) and the background flow, first. A powerful tool is the Stokes number (4) which gives indication of how particles follow the stream lines. It basically indicates how fast a particle can adapt the changes in flow velocities. Stokes number is the ratio of particle relaxation time (τ_p) to fluid flow timescale (τ_{fluid}) decided by agitation rate N in a stirred tank (5). Even by speculating orders of magnitude for particle diameter (d_p), liquid viscosity (μ) and particle density order of magnitude for particle relaxation time can be reasonably predicted (6). In case of microbes in stirred tanks, St << 1 which means that cells can be approximated with massless particles which follow the streamline. A consequence of weightlessness is that randomness generated from turbulence will be undermined. Hence, superimposition by the onset of random walk option should be enabled.

$$St = \frac{\tau_p}{\tau_{\text{fluid}}} \tag{4}$$

$$\tau_{\text{fluid}} = {}^{1}\!/_{N} \tag{5}$$

$$\tau_p = \frac{\rho_p d_p^2}{18\mu_{\text{liquid}}} \tag{6}$$

So far, those settings could be fixed using the already implemented tunings of the software. However, fine tuning of the modeling tasks usually requires further inputs such as specification of microbial kinetics and particle tracking via so-called user-

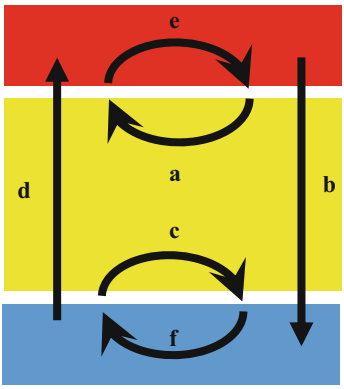


Fig. 5 Trajectories of particles traveling through overflow metabolism (*red*) "O," limitation (*yellow*) "L," and starvation zone (*blue*) "S," (**a**) OLO, (**b**) OLS, (**c**) SLS (**d**) SLO, (**e**) LOL, and (**f**) LSL

defined functions (UDFs). The approach is called "one-way" coupling illustrating that particles do not affect the background flow field which is "frozen" in pseudo-steady state. For this purpose 100,000 particles are introduced into the reactor. Criteria for choosing the number of particles are discussed in the literature [14, 25, 30].

To analyze lifelines with respect to regime shifts, the common approach is to consider the time duration between two consecutive regime transitions as the individual retention time inside one regime. Accordingly, transitions may be categorized by six trajectories as shown in Fig. 5. For example, trajectory "e" accounts for transitions from excess to limitation and back [13].

3.4 Lagrangian Readouts

Once trajectories are sorted, frequencies of retention times may be plotted and analyzed to illustrate their significance. In accordance with the ergodicity criterion, lifelines are recorded after one mixing time and for 10 mixing times thereafter (Fig. 6). A practical approach to reduce computational times is to distribute particles equally inside the bioreactor at the beginning.

As expected, short residence times largely contribute to the residence time distributions. As shortest residence times basically represent artifacts, residence time distributions must be filtered properly before further processing [12, 14]. However, lifelines of Fig. 7 still describe fair estimates of real, nonideal reactors. Their impact on bioprocesses with elevated biomass concentration and/or high cellular metabolic activity is of particular interest for bioprocess analysis.

For instance, one may wonder how simulation readouts may be translated into wet-lab-scale-up simulators that mimic simulations already in lab scale. Figure 8 illustrates some putative settings following suggestions of Kuschel and Takors [26] and Noorman [51]. Basically, the following compartments should be considered:

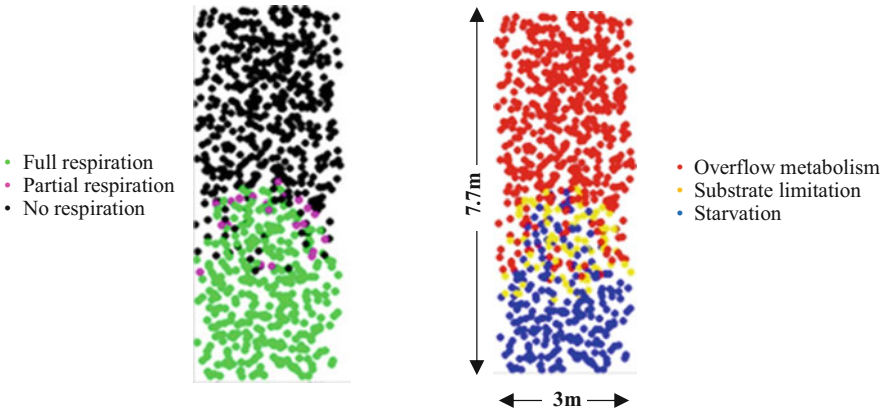


Fig. 6 Particles distribution over the tank volume: oxygen concentration regimes indicating respiration in *green*, partial respiration in *magenta*, anaerobic conditions in *black* (*left*), glucose concentration regimes where overflow metabolism is indicated with *red*, limitation zone with *yellow*, and starvation compartment with *blue* (*right*)

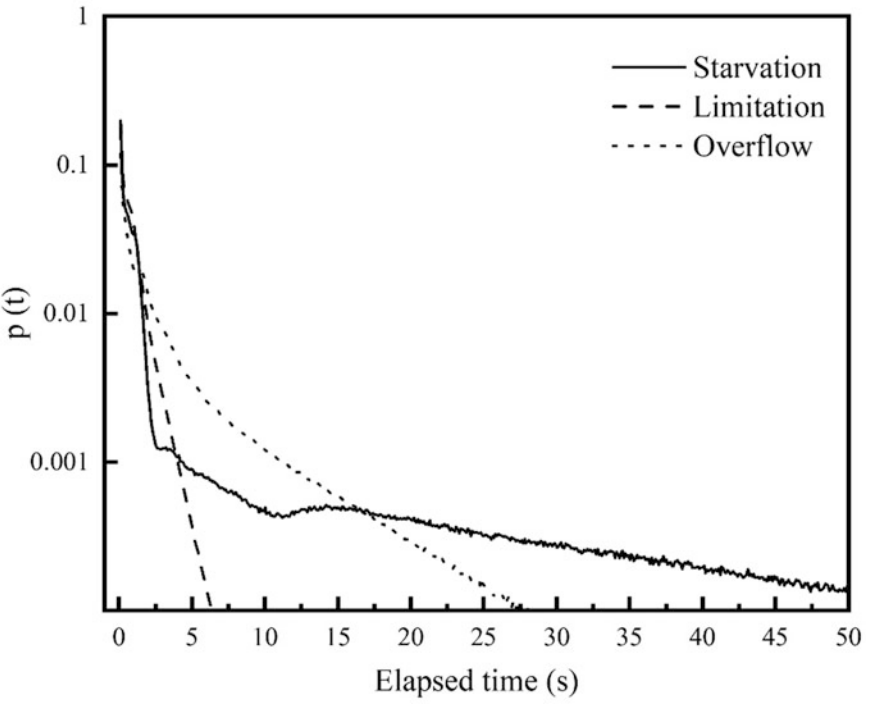


Fig. 7 Residence time distributions for three compartment design

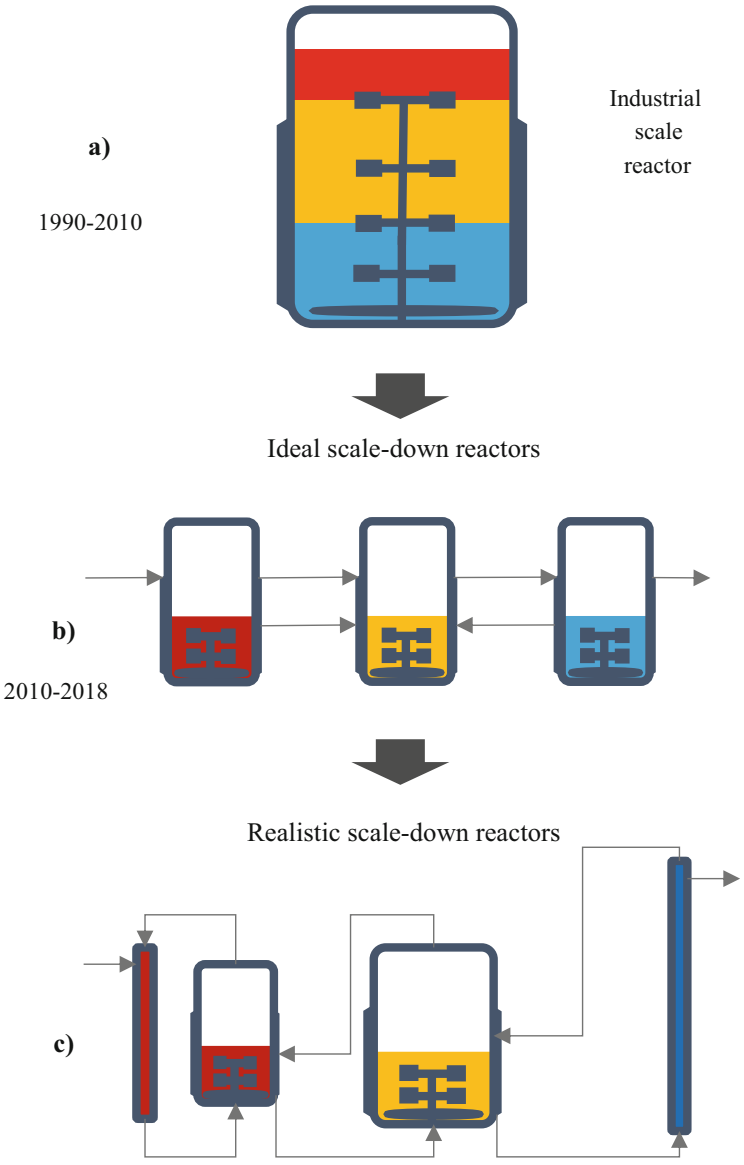


Fig. 8 Evolution of the scale-down bioreactor: (a) detection of the gradients sparked the interest in scale-down bioreactors which in turn led to (b) a conventional approach of series of ideal bioreactors which laid the ground work for the next generation of (c) nonideal but more representative scale-down bioreactor

1. A nonideal stirred tank with a bypass flow to imitate the smaller timescales and dead zones to reproduce the spread of residence time distribution for starvation zone.

2. A stirred tank reactor to represent overflow metabolism compartment.
3. A nonideal stirred tank with a strong bypass flow of a nonideal plug flow reactor with high dispersion to reproduce residence time distribution of limitation zone.

The implementation of such designs may face technical challenges in reality finally causing well-balanced decisions between simulation agreement and hardware constraints. In general, such designs are in line with multi-compartment bioreactor settings already used in the labs [60–62]. Accordingly, the combination with CFD calculations allows a quantitative estimate of how close lab performance mimics large-scale conditions.

4 Scale-Down Examples and Methods from the Literature

As mentioned earlier, the scale-down procedure should be carried out with the end in mind, and this results in various optimization targets based on the scenario of interest. With this in mind, we briefly go through the literature to evaluate different design goals addressed using this approach. One of the first applications of scale-down can be seen from the work of George et al. [63] that was a two-compartment scale-down with one stirred tank reactor (STR) connected to a plug flow reactor (PFR). The impact of the overflow metabolism was investigated and to ensure such condition substrate was fed at PFR entrance. One of the design considerations in such setups is to maintain proper dissolved oxygen (DO) levels in the scale-down reactor to prevent deviating from an industrial case. For this PFR should also be aerated without major distributions in the flow pattern. The same setup was later used to evaluate the effect of substrate heterogeneity of acetate formation [7]. One way to achieve this is to use oxygen gas in PFR compartment to regulate the DO levels. On a side note, other scale-down studies were carried before, but they simply were not titled, so an example of this is the work of Sweere et al. [64] and Sweere et al. [40] which was concerned with DO fluctuations by using STR compartments while investigating the impact of the tubing. Hewitt et al. [59] took interest in imperfections occurring at industrial scale in *E. coli* fermentations with high cell density and STR-PFR scale-down reactor design. Enfors et al. [65] also used an STR-PFR configuration to shed light on by-product formation in yeast for different feeding locations in the scale-down reactor. With a different perspective, Delvigne et al. [66] presented a stochastic approach to reproduce hydrodynamic flow fields in the scale-down reactor based on residence time distribution for four different configurations, namely, two STR-PFR with various PFR diameters, STR-2PFR (parallel) and STR-2PFR (series). High flexibility and attractiveness of scale-down approach resulted in its application in more complicated bioprocesses like algal bioprocesses where fluid flow was simulated by CFD and the illumination was included using the Monte Carlo method [67]. Even with only one or two compartments at hand, a diverse range of designs can be generated from a single vessel with fluctuating feed to an STR-PFR with mixing and aeration in PFR and, hence, provide

access to specific metabolic data in industrial cases [68]. Versatility of scale-down reactors allows them to shed light on new characteristics of the process from repetitive short stimuli faced by cells [17] and cell cycle in a yeast fermentation [27] to mammalian cell cultures with specific interest in strain rate and carbon dioxide gradients [69]. Once it is determined that such systems are not representative of the desired conditions, more complex designs are considered if need be [62].

5 Advantages and Considerations

In this section we briefly go through some of the reasons that made EL simulations relevant in the scaling up and down. We will also discuss some of the key considerations which should be address accordingly.

Compared to other methods that tend to address the question of heterogeneity, EL offers intrinsically superior features compared to other methods. Although noticeable attempts were made to equip CFD simulations with a PBM for different classes of growth rate [21], including a microorganism's history is one of the essential aspects that encourage the use of EL simulations over PBMs. This can serve as a valuable tool in evaluating transient phenomena. Another benefit of using cell tracking is that it allows integrating of multiple variables in growth rate calculations without any major compromise in calculation time [36].

A versatile tool such as EL should be handled delicately. Assumptions from a wide spectrum of phenomena from turbulence to cellular heterogeneity and regulatory networks within the cell have to be modeled with utmost attention to balance the numerical burden with biological accuracy. Since the pioneering work of [36], significant efforts have been made to find a robust foundation for E-L simulations. Here are some of the lessons learned along the way:

- Flow field considerations

Since the start of EL simulations in bioprocess engineering, computational resources have become more abundant and yet still are not enough to complete a fully coupled transient simulation in a matter of hours or days. Having said that it might be the time for the scale-down engineering to revisit some of the basics, that could improve the reality of the simulations [13, 22]. The state-of-the-art EL simulations mainly take place in a stepwise fashion introducing some artifacts to the solution, for instance, when the turbulence from a transient solution is frozen in a snapshot. Accordingly, one may wonder about the validity of lifelines extracted at that distinct time window. Still, even with such compromises, EL is offering valuable insights especially to the heterogeneity within the bioreactor.

- Particle tracking considerations

One of the key points is to consider the coupling mechanisms between Eulerian phases and the Lagrangian particle. Basic assumption is that the condition $St < 0.01$ holds true for a single cell, and hence it can be assumed that the cell itself has no

effect on the flow. Currently the number of particles being tracked is in the order of 100,000–1,000,000 particles. This, on the one hand, forces the investigation to adapt an extra assumption which is to associate a fix biomass to the Eulerian flow field and consider simulating a quasi-steady state [15]. On the other hand, associating the biomass with particles might introduce artificial concentration gradients especially in the spatial locations that are less visited by the limited number of particles. In an ideal scenario, it is desired to have a large enough set of particles that results in a statistically stable gradient although if the number of particles is too large, apart from computational resource limitations, it could be that a cluster of particles will be present at the same cell where the limited number of cells will have access to the substrate and the others will face starvation. This is of utmost importance for coupling consumption to concentrations (two-way coupled metabolism). Among other things, agglomeration is not considered, and the heterogeneities arising from such agglomerations can be an interesting question to tackle by assuming the particle as a swarm center point. Based on the final goal of the simulation, valid assumptions are to be made so that most of the concerning happenings at reasonable numerical/computational costs are revealed.

- Kinetic considerations

The choice of proper kinetic models depends on the traits of the microbe. Particular focus should be given to the timescale of cellular response. Conventional black-box, nonstructured kinetics typically initiate instantaneous microbial responses with respect to substrate uptake, growth, etc. However, the consideration of structured models additionally considering multiple cellular regulation levels integrates delayed responses, most likely. The latter cause a spatial disconnection of initiation and cellular response inside the bioreactor [2, 18]. Besides, even the consideration of proper metabolic models reacting instantaneously on microenvironmental stimuli may be a challenging task. Modeling the cellular behavior by lumped metabolic pools is considered a reasonable compromise to cope with the complexity of metabolism [30, 36, 70]. It might be intriguing to employ data science techniques like reduced order modeling to decide the features that have the most impact on the behavior in the future. Especially in cases where biomass concentration is low enough to create a large limitation zone but larger enough to increase oxygen sensitivity and resulting in unwanted overflow metabolism production [71–73]. One can modify a lumped metabolic model in a way so that it also takes into account the transient behavior of the population. An abstract illustration of such models is shown in Fig. 9. Here, oxygen concentrations above three times the affinity constant of oxygen [74] are assumed to supply enough oxygen to support the utilization of substrate, which equals to maximum oxygen uptake rate of the cell ($q_{O_2} = q_{O_{2,max}}$). In this region overflow metabolism is expected only when the concentration of the substrate exceeds the respiration capacity $\left(q_s \times Y_{O/S} > q_{O_{2,max}}\right)$. This holds true for most of the bioprocesses that operate in a way that guarantees the oxygen demand ($> 30\%$ DO concentration). This assumption requires revisiting once the rheology is changed. With slight adaptions, it can be

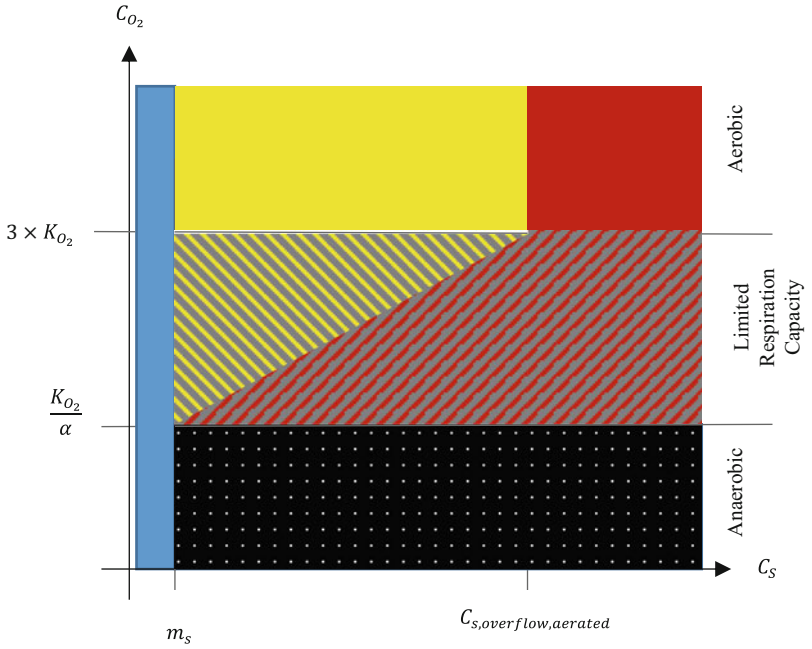


Fig. 9 Regime map for aerated cultures based on respiration capacity. $\mathbf{m_s}$ represents minimum substrate concentration required for cell maintenance, and α is an arbitrary value

assumed that respiration capacity is dependent on oxygen availability itself. As opposed to substrate depletion, some microorganisms still continue to grow in the absence of oxygen. Investigation of oxygen gradients can help shed light on the fluctuations of a single cell experience.

EL simulations have emerged within a bioprocess community to serve as a tool for diagnosis, optimization, and design for both microorganisms and bioreactors alike. It materializes when two digital twins of the bioreactor and microorganism meet and offer new meaning, solutions, and challenges.

6 Conclusion and Outlook

Lifeline studies allow the analysis of large-scale heterogeneities with the eyes of the cells which puts biological criteria into the foreground of strain and process engineering as well as bioreactor design. Computational power is steadily improving allowing the application of said method with ever increasing complexity and less time. Hence it is time to further exploit this tool and making it an integral part of bioprocess and bioreactor design. As such, it may even be used as a digital twin allowing unprecedented insights into cellular needs. More sophisticated simulations

are expected to use digital twins, taking another step toward dynamic simulations which are widely used by the biotech industry and hence pave the way for real-time simulations 1 day. Once these dynamic simulations are validated, it can be used together with powerful tools like augmented reality to develop online monitoring systems once the digital twin of the process is in place.

Acknowledgments This work was partially supported by the German Federal Ministry of Education and Research (BMBF), grant number: FKZ 031B0629.

Abbreviations and Nomenclatures

Abbreviations

CFD	Computational fluid dynamics
DNS	Direct numerical simulation
DO	Dissolved oxygen
EL	Euler-Lagrange
LES	Large eddy simulation
NSE	Navier-Stokes equations
PBM	Population balance model
PFR	Plug flow reactor
RANS	Reynolds-averaged Navier-Stokes
STR	Stirred tank reactor
UDF	User-defined function

Nomenclatures

a	Bubble surface
C_{glucose}	Glucose concentration
C_{O_2}	Dissolved oxygen concentration
$C^*_{O_2}$	Equilibrium oxygen concentration
C_x	Biomass concentration
d_p	Bubble diameter
D	Diffusion coefficient
k	Turbulent kinetic energy
k_l	Mass transfer coefficient
K_{glucose}	Saturation constant for glucose
K_{oxygen}	Saturation constant for oxygen
N	Agitation rate
P	Bioreactor power input
q_s	Specific substrate uptake rate
$q_{s,\,\max}$	Maximum specific substrate uptake rate
q_{O_2}	Specific oxygen uptake rate
$q_{O_2,\max}$	Maximum specific oxygen uptake rate
St	Stokes number
V	Bioreactor volume
$Y_{\frac{x}{s}}$	Biomass yield
$Y_{\frac{o}{s}}$	Oxygen yield
ε	Turbulent kinetic energy dissipation rate
μ	Growth rate
$\mu_{\max}$	Maximum growth rate
μ_{liquid}	Molecular viscosity
ν	Kinematic viscosity

ρ_p	Bubble density
τ_{mix}	Bioreactor mixing time
τ_{fluid}	Fluid timescale
τ_p	Bubble timescale

References

1. Lara AR, Galindo E, Ramírez OT, Palomares LA (2006) Living with heterogeneities in bioreactors: understanding the effects of environmental gradients on cells. Mol Biotechnol 34 (3):355–381. https://doi.org/10.1385/MB:34:3:355
2. Zieringer J, Takors R (2018) In silico prediction of large-scale microbial production performance: constraints for getting proper data-driven models. Comput Struct Biotechnol J 16:246–256. https://doi.org/10.1016/j.csbj.2018.06.002
3. Breuer M, Lakehal D, Rodi W (1995) Flow around a surface mounted cubical obstacle: comparison of les rans-results. In: IMACS/COST conference on CFD, 3D complex flows, Lausanne 1995
4. Larsson G, Törnkvist M, Ståhl Wernersson E, Trägårdh C, Noorman H, Enfors SO (1996) Substrate gradients in bioreactors: origin and consequences. Bioprocess Eng 14(6):281–289. https://doi.org/10.1007/BF00369471
5. Buchholz J, Graf M, Freund A, Busche T, Kalinowski J, Blombach B, Takors R (2014) CO 2 / HCO 3− perturbations of simulated large scale gradients in a scale-down device cause fast transcriptional responses in Corynebacterium glutamicum. Appl Microbiol Biotechnol 98 (20):8563–8572. https://doi.org/10.1007/s00253-014-6014-y
6. Löffler M, Simen JD, Jäger G, Schäferhoff K, Freund A, Takors R (2016) Engineering E. coli for large-scale production – strategies considering ATP expenses and transcriptional responses. Metab Eng 38:73–85. https://doi.org/10.1016/j.ymben.2016.06.008
7. Neubauer P, Häggström L, Enfors S-O (1995) Influence of substrate oscillations on acetate formation and growth yield in Escherichia coli glucose limited fed-batch cultivations. Biotechnol Bioeng 47(2):139–146. https://doi.org/10.1002/bit.260470204
8. Oosterhuis NMG, Kossen NWF (1984) Dissolved oxygen concentration profiles in a production-scale bioreactor. Biotechnol Bioeng 26(5):546–550. https://doi.org/10.1002/bit.260260522
9. Simen JD, Löffler M, Jäger G, Schäferhoff K, Freund A, Matthes J, Müller J et al (2017) Transcriptional response of Escherichia coli to ammonia and glucose fluctuations. Microb Biotechnol 10(4):858–872. https://doi.org/10.1111/1751-7915.12713
10. Teleki A, Sánchez-Kopper A, Takors R (2015) Alkaline conditions in hydrophilic interaction liquid chromatography for intracellular metabolite quantification using tandem mass spectrometry. Anal Biochem 475:4–13. https://doi.org/10.1016/j.ab.2015.01.002
11. von Wulffen J, Ulmer A, Jäger G, Sawodny O, Feuer R (2017) Rapid sampling of Escherichia coli after changing oxygen conditions reveals transcriptional dynamics. Genes 8(3). https://doi.org/10.3390/genes8030090
12. Haringa C, Deshmukh AT, Mudde RF, Noorman HJ (2017a) Euler-Lagrange analysis towards representative down-scaling of a 22 M3 Aerobic S. cerevisiae fermentation. Chem Eng Sci 170:653–669. https://doi.org/10.1016/j.ces.2017.01.014
13. Haringa C, Mudde RF, Noorman HJ (2018a) From industrial fermentor to CFD-guided downscaling: what have we learned? Biochem Eng J 140(April):57–71. https://doi.org/10.1016/j.bej.2018.09.001
14. Haringa C, Noorman HJ, Mudde RF (2017b) Lagrangian modeling of hydrodynamic–kinetic interactions in (bio)chemical reactors: practical implementation and setup guidelines. Chem Eng Sci 157:159–168. https://doi.org/10.1016/j.ces.2016.07.031

15. Kuschel M, Siebler F, Takors R (2017) Lagrangian trajectories to predict the formation of population heterogeneity in large-scale bioreactors. Bioengineering 4(4):27. https://doi.org/10.3390/bioengineering4020027
16. Wang G, Haringa C, Tang W, Noorman H, Chu J, Zhuang Y, Zhang S (2020) Coupled metabolic-hydrodynamic modeling enabling rational scale-up of industrial bioprocesses. Biotechnol Bioeng 117(3):844–867. https://doi.org/10.1002/bit.27243
17. Nieß A, Löffler M, Simen JD, Takors R (2017) Repetitive short-term stimuli imposed in poor mixing zones induce long-term adaptation of E. Coli cultures in large-scale bioreactors: experimental evidence and mathematical model. Front Microbiol 8(Jun):1–9. https://doi.org/10.3389/fmicb.2017.01195
18. Zieringer J, Takors R (2020) Data-driven in-silico prediction of regulation heterogeneity and ATP demands of Escherichia coli in large-scale bioreactors
19. Morchain J, Gabelle JC, Cockx A (2014) A coupled population balance model and CFD approach for the simulation of mixing issues in lab-scale and industrial bioreactors. AICHE J 60(1):27–40. https://doi.org/10.1002/aic.14238
20. Morchain J, Pigou M, Lebaz N (2017) A population balance model for bioreactors combining interdivision time distributions and micromixing concepts. Biochem Eng J 126:135–145. https://doi.org/10.1016/j.bej.2016.09.005
21. Pigou M, Morchain JÔ (2015) Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models. Chem Eng Sci 126(April):267–282. https://doi.org/10.1016/j.ces.2014.11.035
22. Haringa C, Tang W, Wang G, Deshmukh AT, van Winden WA, Chu J, van Gulik WM, Heijnen JJ, Mudde RF, Noorman HJ (2018b) Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: towards rational scale-down and design optimization. Chem Eng Sci 175:12–24. https://doi.org/10.1016/j.ces.2017.09.020
23. Haringa C, Vandewijer R, Mudde RF (2018c) Inter-compartment interaction in multi-impeller mixing: part i. experiments and multiple reference frame CFD. Chem Eng Res Des 136 (June):870–885. https://doi.org/10.1016/j.cherd.2018.06.005
24. Haringa C, Vandewijer R, Mudde RF (2018d) Inter-compartment interaction in multi-impeller mixing. Part ii. Experiments, sliding mesh and large eddy simulations. Chem Eng Res Des 136 (June):886–899. https://doi.org/10.1016/j.cherd.2018.06.007
25. Siebler F, Lapin A, Hermann M, Takors R (2019) The impact of CO gradients on C. ljungdahlii in a 125 m3 bubble column: mass transfer, circulation time and lifeline analysis. Chem Eng Sci 207:410–423. https://doi.org/10.1016/j.ces.2019.06.018
26. Kuschel M, Takors R (2020) Simulated oxygen and glucose gradients as a prerequisite for predicting industrial scale performance a priori
27. Heins AL, Fernandes RL, Gernaey KV, Lantz AE (2015) Experimental and in silico investigation of population heterogeneity in continuous Saccharomyces cerevisiae scale-down fermentation in a two-compartment setup. J Chem Technol Biotechnol 90(2):324–340. https://doi.org/10.1002/jctb.4532
28. Wang T, Wang J, Jin Y (2005) Population balance model for gas - liquid flows: influence of bubble coalescence and breakup models. Ind Eng Chem Res 44(19):7540–7549. https://doi.org/10.1021/ie0489002
29. Venneker BCH, Derksen JJ, Van den Akker HEA (2002) Population balance modeling of aerated stirred vessels based on CFD. AICHE J 48(4):673–685. https://doi.org/10.1002/aic.690480404
30. Lapin A, Schmid J, Reuss M (2006) Modeling the dynamics of E. coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor-A structured-segregated approach. Chem Eng Sci 61(14):4783–4797. https://doi.org/10.1016/j.ces.2006.03.003
31. Dehbi A (2008) A CFD model for particle dispersion in turbulent boundary layer flows. Nucl Eng Des 238(3):707–715. https://doi.org/10.1016/j.nucengdes.2007.02.055

32. Haringa C, Tang W, Deshmukh AT, Xia J, Reuss M, Heijnen JJ, Mudde RF, Noorman HJ (2016) Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: an analysis of organism lifelines. Eng Life Sci 16(7):652–663. https://doi.org/10.1002/elsc.201600061
33. Liu Y, Wang ZJ, Xia JY, Haringa C, Liu YP, Chu J, Zhuang YP, Zhang SL (2016) Application of Euler–Lagrange CFD for quantitative evaluating the effect of shear force on Carthamus tinctorius L. cell in a stirred tank bioreactor. Biochem Eng J 114:209–217. https://doi.org/10.1016/j.bej.2016.07.006
34. Gunyol O, Mudde RF (2009) Computational study of hydrodynamics of a standard stirred tank reactor and a large-scale multi-impeller fermenter. Int J Multiscale Comput Eng:559–576. https://doi.org/10.1615/IntJMultCompEng.v7.i6.60
35. Coroneo M, Montante G, Paglianti A, Magelli F (2011) CFD prediction of fluid flow and mixing in stirred tanks: numerical issues about the RANS simulations. Comput Chem Eng 35 (10):1959–1968. https://doi.org/10.1016/j.compchemeng.2010.12.007
36. Lapin A, Müller D, Reuss M (2004) Dynamic behavior of microbial populations in stirred bioreactors simulated with Euler-Lagrange methods: traveling along the lifelines of single cells. Ind Eng Chem Res 43(16):4647–4656. https://doi.org/10.1021/ie030786k
37. Ducci A, Yianneskis M (2005) Direct determination of energy dissipation in stirred vessels with two-point LDA. AICHE J 51(8):2133–2149. https://doi.org/10.1002/aic.10468
38. Chaouat B (2017) The state of the art of hybrid RANS/LES modeling for the simulation of turbulent flows. Flow Turbul Combust 99(2):279–327. https://doi.org/10.1007/s10494-017-9828-8
39. Fröhlich J, von Terzi D (2008) Hybrid LES/RANS methods for the simulation of turbulent flows. Prog Aerosp Sci 44(5):349–377. https://doi.org/10.1016/j.paerosci.2008.05.001
40. Sweere APJ, Janse L, Luyben KCAM, Kossen NWF (1988a) Experimental simulation of oxygen profiles and their influence on Baker's yeast production: II. Two-fermentor system. Biotechnol Bioeng 31(6):579–586. https://doi.org/10.1002/bit.260310610
41. Sweere APJ, Giesselbach J, Barendse R, de Krieger R, Honderd G, Luyben KCAM (1988c) Modelling the dynamic behaviour of Saccharomyces cerevisiae and its application in control experiments. Appl Microbiol Biotechnol 28(2):116–127. https://doi.org/10.1007/BF00694298
42. Sweere APJ, Matla YA, Zandvliet J, Ch K, Luyben AM, Kossen NWF (1988d) Experimental simulation of glucose fluctuations - the influence of continually changing glucose concentrations on the fed-batch Baker's yeast production. Appl Microbiol Biotechnol 28(2):109–115. https://doi.org/10.1007/BF00694297
43. Pham HTB, Larsson G, Enfors SO (1998) Growth and energy metabolism in aerobic fed-batch cultures of Saccharomyces cerevisiae: simulation and model verification. Biotechnol Bioeng 60 (4):474–482. https://doi.org/10.1002/(SICI)1097-0290(19981120)60:4<474::AID-BIT9>3.0.CO;2-J
44. Serio M, Di RT, Santacesaria E (2001) A kinetic and mass transfer model to simulate the growth of Baker's yeast in industrial bioreactors. Chem Eng J 82(1–3):347–354. https://doi.org/10.1016/S1385-8947(00)00353-3
45. Wright MR, Bach C, Gernaey KV, Krühne U (2018) Investigation of the effect of uncertain growth kinetics on a CFD based model for the growth of S. cerevisiae in an industrial bioreactor. Chem Eng Res Des 140:12–22. https://doi.org/10.1016/j.cherd.2018.09.040
46. Sokolichin A, Eigenberger G, Lapin A, Lübbert A (1997) Dynamic numerical simulation of gas-liquid two-phase flows: Euler/Euler versus Euler/Lagrange. Chem Eng Sci 52(4):611–626. https://doi.org/10.1016/S0009-2509(96)00425-3
47. Ireland PJ, Desjardins O (2017) Improving particle drag predictions in Euler–Lagrange simulations with two-way coupling. J Comput Phys 338:405–430. https://doi.org/10.1016/j.jcp.2017.02.070
48. Linkès M, Fede P, Morchain JÔ, Schmitz P (2014) Numerical investigation of subgrid mixing effects on the calculation of biological reaction rates. Chem Eng Sci 116:473–485. https://doi.org/10.1016/j.ces.2014.05.005

49. Löffler M, Simen JD, Müller J, Jäger G, Laghrami S, Schäferhoff K, Freund A, Takors R (2017) Switching between nitrogen and glucose limitation: unraveling transcriptional dynamics in Escherichia coli. J Biotechnol 258(April):2–12. https://doi.org/10.1016/j.jbiotec.2017.04.011
50. Liné A, Gabelle JC, Morchain J, Anne-Archard D, Augier F (2013) On POD analysis of PIV measurements applied to mixing in a stirred vessel with a shear thinning fluid. Chem Eng Res Des 91(11):2073–2083. https://doi.org/10.1016/j.cherd.2013.05.002
51. Noorman H (2011) An industrial perspective on bioreactor scale-down: what we can learn from combined large-scale bioprocess and model fluid studies. Biotechnol J 6(8):934–943. https://doi.org/10.1002/biot.201000406
52. Clift R, Grace JR, Weber ME (2005) Bubbles, drops, and particles. In: Clift R, Grace JR, Weber ME (eds) Dover books on engineering. Dover, Mineola
53. Lamont JC, Scott DS (1970) An eddy cell model of mass transfer into the surface of a turbulent liquid. AICHE J 16(4):513–519. https://doi.org/10.1002/aic.690160403
54. Roels JA (1983) Roels JA (ed) Energetics and kinetics in biotechnology. Elsevier Biomedical Press, Amsterdam
55. Senn H, Lendenmann U, Snozzi M, Hamer G, Egli T (1994) Biochi ~ Mic ~ a et biophysica A ~ Ta the growth of Escherichia coli in glucose-limited chemostat cultures: a re-examination of the kinetics. Sci Technol 1201:424–436
56. Kita K, Konishi K, Anraku Y (1984) Terminal oxidases of Escherichia coli aerobic respiratory chain. J Biol Chem 259(5):3368–3374
57. Valgepea K, Adamberg K, Vilu R (2011) Decrease of energy spilling in Escherichia coli continuous cultures with rising specific growth rate and carbon wasting. BMC Syst Biol 5. https://doi.org/10.1186/1752-0509-5-106
58. Jain R, Srivastava R (2009) Metabolic investigation of host/pathogen interaction using MS2-infected Escherichia coli. BMC Syst Biol 3. https://doi.org/10.1186/1752-0509-3-121
59. Hewitt CJ, Von Caron GN, Axelsson B, McFarlane CM, Nienow AW (2000) Studies related to the scale-up of high-cell-density E. coli fed-batch fermentations using multiparameter flow cytometry: effect of a changing microenvironment with respect to glucose and dissolved oxygen concentration. Biotechnol Bioeng 70(4):381–390. https://doi.org/10.1002/1097-0290(20001120)70:4<381::AID-BIT3>3.0.CO;2-0
60. Takors R (2012) Scale-up of microbial processes: impacts, tools and open questions. J Biotechnol 160(1–2):3–9. https://doi.org/10.1016/j.jbiotec.2011.12.010
61. Ankenbauer A, Schäfer RA, Viegas SC, Pobre V, Voß B, Arraiano CM, Takors R (2020) Pseudomonas putida KT2440 is naturally endowed to withstand industrial-scale stress conditions. Microb Biotechnol:635536. https://doi.org/10.1111/1751-7915.13571
62. Delvigne F, Takors R, Mudde R, van Gulik W, Noorman H (2017) Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. Microb Biotechnol 10(5):1267–1274. https://doi.org/10.1111/1751-7915.12803
63. George S, Larsson G, Enfors SO (1993) A scale-down two-compartment reactor with controlled substrate oscillations: metabolic response of Saccharomyces cerevisiae. Bioprocess Eng 9(6):249–257. https://doi.org/10.1007/BF01061530
64. Sweere APJ, Mesters JR, Janse L, Luyben KCAM, Kossen NWF (1988b) Experimental simulation of oxygen profiles and their influence on Baker's yeast production: I. One-fermentor system. Biotechnol Bioeng 31(6):567–578. https://doi.org/10.1002/bit.260310609
65. Enfors SO, Jahic M, Rozkov A, Xu B, Hecker M, Jürgen B, Krüger E et al (2001) Physiological responses to mixing in large scale bioreactors. J Biotechnol 85(2):175–185. https://doi.org/10.1016/S0168-1656(00)00365-5
66. Delvigne F, Lejeune A, Destain J, Thonart P (2006) Stochastic models to study the impact of mixing on a fed-batch culture of Saccharomyces cerevisiae. Biotechnol Prog 22(1):259–269. https://doi.org/10.1021/bp050255m
67. Sastre R, Rosa ZC, Perner-Nochta I, Fleck-Schneider P, Posten C (2007) Scale-down of microalgae cultivations in tubular photo-bioreactors-A conceptual approach. J Biotechnol 132(2):127–133. https://doi.org/10.1016/j.jbiotec.2007.04.022

68. Neubauer P, Junne S (2010) Scale-down simulators for metabolic analysis of large-scale bioprocesses. Curr Opin Biotechnol 21(1):114–121. https://doi.org/10.1016/j.copbio.2010.02.001
69. Paul K, Herwig C (2020) Scale-down simulators for mammalian cell culture as tools to access the impact of inhomogeneities occurring in large-scale bioreactors. Eng Life Sci 20 (5–6):197–204. https://doi.org/10.1002/elsc.201900162
70. Tang W, Deshmukh AT, Haringa C, Wang G, van Gulik W, van Winden W, Reuss M et al (2017) A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by Penicillium chrysogenum. Biotechnol Bioeng 114 (8):1733–1743. https://doi.org/10.1002/bit.26294
71. Lei F, Rotboll M, Jorgensen SB (2001) A biochemically structured model for Saccharomyces cerevisiae. J Biotechnol 88(3):205–221. https://doi.org/10.1016/S0168-1656(01)00269-3
72. Rizzi M, Baltes M, Theobald U, Reuss M (1997) In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical model. Biotechnol Bioeng 55(4):592–608. https://doi.org/10.1002/(SICI)1097-0290(19970820)55:4<592::AID-BIT2>3.0.CO;2-C
73. Vanrolleghem PA, De Jong-Gubbels P, Van Gulik WM, Pronk JT, Van Dijken JP, Heijnen S (1996) Validation of a metabolic network for Saccharomyces cerevisiae using mixed substrate studies. Biotechnol Prog 12(4):434–448. https://doi.org/10.1021/bp960022i
74. Bailey J, Bailey JE, Ollis DF, Simpson RJ, Ollis DF (1986) Biochemical engineering fundamentals. McGraw-Hill chemical engineering series. McGraw-Hill. https://books.google.de/books?id=KM9TAAAAMAAJ