Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Engineering

Bioprocess systems analysis, modeling, estimation, and control

Yu Luo, Varghese Kurian and Babatunde A Ogunnaike

The production of monoclonal antibody (mAb) therapeutics, a rapidly growing multi-billion-dollar enterprise in the biopharmaceutical industry, faces major challenges in achieving desired productivity and product quality consistently. These challenges, traditionally addressed by genetic engineering and media recipe development, are now being addressed with process systems engineering (PSE) techniques. In this perspective paper, we discuss how this alternative approach, comprising three components — process modeling, estimation, and control — is being used to address biomanufacturing challenges. We survey the state of current practice for each component, identify existing gaps, and highlight some advances needed to achieve routine implementation of fully automated systems for optimal bioprocess operations.

Address
Chemical and Biomolecular Engineering, University of Delaware, Newark, DE, United States

Corresponding author: Ogunnaike, Babatunde A (ogunnaike@udel.edu)

Current Opinion in Chemical Engineering 2021, **33**:xx-yy

This review comes from a themed issue on **Biotechnology and bioprocess eng: mechanistic and data-driven modelling of Bioprocesses**

Edited by **Cleo Kontoravdi** and **Colin Clarke**

https://doi.org/10.1016/j.coche.2021.100705

Introduction

Valued at USD 239.8 billion in 2019 and forecasted to grow at an annual rate of over 13% [1] — with therapeutic monoclonal antibodies (mAbs) alone expected to generate a revenue of USD 300 billion by 2025 [2] — the biopharmaceutical industry faces major challenges in achieving desired productivity and product quality consistently [3]. The industry has traditionally addressed these challenges by focusing on improving the manufacturing process via such techniques as genetic engineering and media recipe development. Recently, interest has grown in a process systems engineering (PSE)-based approach, whereby the challenges are addressed via process modeling, estimation, and control. This paper is concerned with reviewing the current status and future needs of the PSE-based approach to meeting the productivity *and* product quality goals in biomanufacturing simultaneously, efficiently, and consistently. The three main components (i) developing mathematical models (ii) estimating the system states and (iii) implementing model-based optimal control strategies — are presented in turn, discussing the state of current practice, identifying existing gaps, and highlighting some of the advancements that will be needed to achieve routine implementation of fully automated systems for optimal bioprocess operations in industrial practice.

Overview and scope

Even though the discussion in this paper applies to bioprocesses in general, our focus will be on the manufacture of mAbs using fed-batch mammalian cell cultures, as an exemplar. The primary operational objective of *all* bioprocesses is ensuring both high productivity and high product quality consistently. Specifically for mAb manufacturing, the most common productivity attributes are protein titer and biomass, and the critical quality attributes (CQAs) of interest are extent of glycosylation, glycan distributions, charge variants, and protein aggregation [4]. In particular, glycosylation, a post-translational modification where glycans are added to the protein, is known to affect product attributes including stability, *in vivo* efficacy, and cytotoxicity [4,5]. As a result of its importance, significant effort has been devoted to understanding *what* affects the cellular process of glycosylation (e.g. process conditions, media supplementation, cell lines genetics [5]), and *how* these factors influence it. *The scope of this paper is therefore limited to modeling, estimation, and control of glycosylation in the manufacture of mAbs.*

The rest of the paper is organized as follows: we discuss modeling first in Section 'Modeling', estimation next in Section 'Estimation', and control last in Section 'Control', with conclusions in Section 'Conclusions'. All sections are organized identically:

- First, we define each topic (modeling, estimation, control) within the context of bioprocesses, with emphasis on glycosylation.
- Next, we discuss the role of the specific section's topic in enabling us to meet the objectives of bioprocess operations.
- Finally, we discuss the current status of the topic and present our opinions about future needs (see Table 1 for a list of cited references for each topic).

Table 1

List of cited references

Main category	Sub category	Reference
Modeling	Review	[6]
	Kinetic-based	[7–14]
	Flux-based	[15–19]
	Data-driven	[20–23]
	Design of experiments	[24–27]
	Hybrid	[28,29•,30•]
	Agent-based	[31–33]
	Automatic generation	[34,35]
Estimation	On-line cell culture sensing	[36–38]
	On-line glycosylation sensing	[39–43]
	Soft sensors (non-bioprocess)	[44–47]
	Soft sensors (bioprocess)	[48,49•,50]
Control	Nutrient/metabolite	[51,52•,53,54]
	Product attribute (open loop)	[55,56•,57,58]
	Product attribute (closed loop)	[59,60•]
	Software–hardware integration	[61]

Modeling

What and why

What is modeling? In the context of bioprocess systems analysis, modeling is the process by which one develops — and validates — mathematical representations of a bioprocess. The resulting models take the form of algebraic, ordinary differential, or partial differential equations, or combinations thereof; they can be obtained on the basis of first-principles mechanisms, or strictly from data, or from a combination of the two.

Why modeling? A validated bioprocess model is useful for at least three things:

1. *Process understanding:*
   (a) Every experiment performed on a process can only generate a specific answer to the specific question asked of the process by the experimental design. For mAb manufacturing processes, where experiments are lengthy and costly, a validated, high-fidelity model allows us to investigate process behavior under various conditions, rapidly and inexpensively. The model also allows general analyses of fundamental process characteristics not possible with individual experiments. For example, in [58], a validated theoretical model was used to present the concept of 'output controllability' of glycosylation — the extent to which the process of glycosylation can be controlled — an intrinsic characteristic that is impossible to deduce from data.
   (b) When the process does not yet exist (for example, during the process design phase), a model can serve as a convenient surrogate for generating simulation data.
   (c) Mathematical models can also be used to design optimal experiments that will produce high information content data with minimum expenditure of experimental effort.
2. *Process variable and output prediction:* Models can be used to estimate the values of internal process variables that are not measurable on-line (e.g. metabolic fluxes within the cell, glycan distribution on the mAb protein molecule), and for predicting important process outputs and attributes, such as productivity and product quality, that are either also unmeasurable on-line, or can only be measured infrequently.
3. *Process control:* A model that captures the relationship between the set of manipulated variables and the set of controlled variables of a process is central to the implementation of model-based control. Such models can be (and have been) used successfully to determine the changes in the manipulated variables necessary to achieve the desired objectives for the controlled variables.

Current status

Existing cell culture models for predicting cell growth, metabolism, and mAb production, and mAb glycosylation models, generally fall into one of two broad categories: mechanistic and data-driven [6]. The fundamental structure and basis of each model type determines its limitations and the application for which it is most appropriate.

Mechanistic models are based on first-principle mechanisms underlying the process in question. By definition, developing such models requires extensive domain knowledge and a substantial amount of effort. Nevertheless, even purely mechanistic based models rely on data for parameter estimation and for model validation. From the statistical theory of parameter estimation, the data requirement for estimating model parameters with acceptable precision varies in proportion to the number of parameters in question. In general, fully mechanistic models usually contain a large number of parameters, some of which may not be determinable independently with any degree of confidence. Under these circumstances, it is customary to adopt 'semi-mechanistic' approaches where only subsets of the known mechanisms are included, replacing the rest with appropriate empirical approximations, resulting in models with fewer parameters [59,7]. Mechanistic models can only predict what the mechanisms incorporated into the model allow. Often, for mAb manufacturing processes, the mechanism of how external inputs such as media recipes, supplements, operating conditions, etc., affect bioprocess dynamics is not understood sufficiently well, and such external inputs are therefore not included in a mechanistic model. These are the 'known unknowns'; we know they exist, but their mechanisms of operation are largely unknown. Adapting a mechanistic model for new cell lines and/or new process conditions to account for the effects these 'known unknowns' requires structural modifications to the model.

A brief discussion of such models and how to adapt them efficiently can be found in [30•].

Mechanistic models may be divided further into kinetics-based models and flux-based models. Kinetics-based models of cell culture typically consist of systems of ordinary differential equations (ODEs), where cell metabolism dynamics are described by variations of the Monod equation of cell growth [8–10]. On the other hand, kinetics-based models of glycosylation typically consist of systems of partial differential equations (PDEs), in order to capture adequately, both temporal and spatial dynamics of the glycosylation reactions taking place in the Golgi apparatus [11,12]. These kinetics-based models are most appropriate for explaining observed phenomena, making predictions of process behavior, and for analyzing intrinsic bioprocess characteristics such as controllability [58]. They are usually not appropriate for situations where the model needs frequent updates.

Flux-based models are typically used to describe steady-state, genome-level behavior of cellular metabolism [15–19]; they are not suitable for dynamic control. It is technically possible to construct dynamic flux-based models by introducing dynamic uptake rates of nutrients as the model inputs; however, these inputs cannot be used for process control because they cannot be manipulated directly. Flux-based models are most appropriate for cell-line development where genome-level characteristics of the cells are altered to achieve certain desired process behavior.

Data-driven models, unlike mechanistic models, describe the relationships among process inputs and process outputs, strictly from experimental data, using equations that have no biological basis. The parameters of such empirical models have no direct connection to biological mechanisms and are estimated entirely from data (e.g. [23]), making these models merely *descriptive* of experimental observation, not *explanatory*. Nevertheless, this modeling paradigm is useful under many conditions of practical importance.

First, when not much is known about the fundamental mechanisms underlying the process in question, one can always employ judiciously-designed experiments to obtain process data, from which one can then develop empirical models for various applications, including control systems design. Also, data-driven models are often able to capture complex dynamics effectively using relatively simple model structures. Even when sufficient mechanistic information is available, the intended use of the model may require that we choose a simpler-structured, data-driven approach. For example, process models intended for controller design need not be mechanistic; in fact, under certain circumstances, a mechanistic model might be too complex for on-line control applications [11,13]. Also, when the objective is to determine the design space for an existing pharmaceutical process, it is often sufficient — sometimes even *preferred* — to employ strategies of statistical design of experiments and the resulting ANOVA/simple regression models [24–26]. In terms of model development, the data-driven approach enjoys several advantages, including the availability of many modeling technique options, for example, basic linear/nonlinear regression, principal component/partial least squares regression (PCR/PLSR), markov chains, artificial neural nets [15,20–23]. In addition, because the model structure is not restricted by the underlying biological mechanisms, the data-driven approach can be used across different processes, and the resulting models are relatively easy to adapt for various process conditions, given that considerable effort is involved in data generation and model validation. However, since the parameters of data-driven models usually have no physical significance, this class of models cannot be used to *explain* process behavior — a potential impediment to regulatory approval.

Developing data-driven models is not feasible when the process in question does not yet exist, or more than likely, when the process cannot generate enough appropriate data because of prohibitive cost and/or effort.

Less-common models such as agent-based models (ABMs) [31,32] have been developed for bioprocesses. ABMs employ artificial and autonomous agents to represent components of the process, with each agent programmed to follow a set of rules to update its state and respond to the environment. This approach gives rise to a realization of the emergent system-level behavior, which may otherwise be analytically intractable within the traditional explicit equation-based modeling framework. Used widely in the social sciences, agent-based modeling is still relatively new to molecular systems biology [33].

While not as easily amenable to theoretical analyses as mechanistic models, ABMs are nevertheless useful for systems whose components naturally admit of representation as 'agents'. For instance, the process of glycosylation could potentially benefit from this approach because glycans, enzymes, and nucleotide sugar donors can all be represented naturally as dynamic agents. The ABM framework therefore offers an intuitive, low-parameterization, and readily adaptive modeling alternative to the current first-principle continuum reaction kinetics approach, for modeling the non-continuum phenomenon of low number reactions taking place in spatially confined regions of the Golgi apparatus. What is lost in the generality and analytical insight accruing from first-principle models may be more than compensated for by the more accurate and more flexible representation of the ABM.

Future needs

Hybrid modeling. Data-based models have the twin advantage of convenience and versatility demonstrated in other applications outside of bioprocessing, but are not useful for mechanistic understanding. Mechanistic models excel in providing mechanistic insight, but are structurally inflexible. Such complementarity argues for a *hybrid* approach to model development, which involves a judicious combination of mechanistic knowledge and data, offering the potential to use one approach to compensate for the weakness of the other. The primary challenge is knowing how best to combine the two approaches appropriately for any specific application, but hybrid modeling of bioprecesses is emerging and will likely fill an appropriate modeling space in industrial applications in the near future. Different applications of this approach in practice may be found in [28,29•,30•].

Design of experiments. Developing any useful bioprocess model requires high-quality, information-rich data. Statistical design of experiments (DoE) methods help to avoid investing significant time and effort into the performance of bioprocess experiments only to generate data sets that are not very informative. The efficiency with which information-rich data sets are generated and analyzed have made DoE methods popular in many scientific endeavors. Of particular interest are *optimal experimental designs (OEDs)* whose explicit objective is to specify experimental conditions required to generate data that will minimize (or maximize) some desired quantity of interest, such as the precision of the process parameter estimates. These techniques are still emerging as useful tools in the biopharmaceutical industry [27].

Multiscale modeling. Another area with potential for strong impact in the future is efficient multiscale modeling [14]. Bioprocesses naturally involve subprocesses occurring at multiple time- and length-scales. At the macro-scale are the activities occurring in the bioreactor, manifested as changes in such bulk reactor characeristics as nutrient concentrations and cell density. These can be represented quite well by assuming that the bioreactor content is well-mixed. The metabolic activities taking place within the cell occur at a length scale much smaller than that within the bulk of the bioreactor. At a still smaller scale are the enzymatic glycosylation reactions taking place within the Golgi apparatus, an organelle within the cell. Consequently, a model that purports to capture important details of the activities in a bioreactor must involve a micro-scale model of the Golgi apparatus combined with a meso-scale model of cell metabolism, connected to a macro-scale model of the bioreactor bulk. Furthermore, regardless of length scale, some reactions occur at a much faster rate than others, so that such models must also represent at least two time-scales: slow and fast.

Such high-fidelity multiscale models are useful for creating bioprocess surrogates (digital twins) for technology transfer from bench to production. However, capturing dynamics at every possible scale makes the resulting model potentially too complex and unwieldy. The challenge of developing computationally tractable, high-fidelity multi-scale models will need to be met effectively before such models can become fully accepted.

Automated model generation. What if there exists a means by which one can automate bioprocess model generation? If achievable, this could greatly facilitate bioprocess development in general. An automated model generation system will significantly lower — if not completely eliminate — the most daunting barrier to widespread use of models in the biopharmaceutical industry: the amount of time, effort, and resources — financial and human — required to build, validate, and deploy bioprocess models. While such a system does not currently exist, some desirable component modes would include:

- *Automated model selection*: from a library of model templates, based on user-specified criteria, which may include the intended use of the model, the nature of available process information, etc.
- *Automated data generation*: from appropriate (optimal) experimental design, and also performing the experiments (via robotics?), collecting the data, and incorporating the results into the model, *automatically*.
- *Data pre-processing*: to remove outliers, impute missing values, smooth noisy signals, and reduce the dimensionality of the data.
- *Symbolic Regression* (or similar techniques): to obtain appropriate parameters fit to data, given a class of potential models with corresponding 'built-in' mathematical expressions.

With on-going rapid development of artificial intelligence (AI) technology, such a dream might become reality sooner rather than later [34,35].

Estimation

What and why

The term 'estimation' (or *state estimation*) in this context means specifically the inference of the complete set of process *state variables* based on available measurements. Many bioprocess variables are either not available for measurement at all, or can only be measured infrequently, through off-line laboratory analyses. For example, in the manufacture of mAbs, biomass, and such product quality attributes as glycan distribution, can only be determined off-line. Yet, these unmeasured variables are often the attributes upon which the acceptability of the manufactured product in end-use is based.

While we may not be able to measure all process variables of interest directly and as frequently as desired, estimating their values accurately is critical to safe, effective, and economical bioprocess operation. In the context of mAb manufacturing, the fundamental premise of state estimation may be understood as follows:

> The final product attributes of a mAb, which are being determined during the manufacturing process, and which can only be measured at run end, depend on affiliated variables (nutrient concentrations, dissolved oxygen (DO) level, etc.), which *can* be measured on-line; these latter variables therefore encode information about the unmeasured, but critical to quality variables, *if only we knew how to 'extract' such information*.

State estimation is the technique for inferring the values of these critical, but unmeasured quantities using available measurements.

Current status

On-line measurement acquisition for off-line measurable quantities. Protein and metabolite concentrations, biomass, and product quality attributes in mAb manufacturing are currently determined via off-line analysis of infrequent bioprocess samples, using such analytical techniques as liquid chromatography (LC), mass spectrometry (MS), capillary electrophoresis, etc. While such measurements, which are relatively accurate and precise, are appropriate for statistical analysis and model development, low sampling frequency, long turnaround time, and high operating costs, make them unsuitable for feedback control. The loss of material due to sampling constitutes yet another disadvantage of these off-line measurement techniques.

The only bioprocess measurements that are routinely available on-line and relatively frequently, are of such bioreactor conditions as DO, pH, and temperature. With the advent of process analytical technologies (PATs), there is now strong motivation for the development of novel techniques for real time measurement of most process variables. For instance, Raman spectroscopy, infrared (IR) spectroscopy, and capacitance probes have recently been deployed successfully for acquiring cell culture measurements in real time [36,37]. Glycosylation site occupancy, a CQA, has also been successfully monitored using *in situ* Raman spectroscopy [38].

Detecting and characterizing glycans at the molecular level present far more formidable challenges than measuring aggregate cell culture quantities. Currently, the most reliable methods for measuring glycan distributions are still LC-based, MS-based, or a combination of both. Several attempts have been made to automate glycan profiling, with acceptable turnaround times. For instance, an automated robotic platform for purification and analysis of mAbs taken directly from bioreactor was developed by Doherty *et al.* [39]. A high-throughput, automated platform based on LC–MS to monitor multiple glycan attributes was developed by Dong *et al.* [40], and another for monitoring multiple CQAs by Chi *et al.* [41]. Lectin microarrays are emerging as a potential alternative to the LC-MS-based techniques. These lectin-based sensors can measure the glycan concentrations directly without having to digest the proteins or release the glycans, and have high sensitivity toward glycan variations [42]. However, the technique is still maturing and has yet to overcome such limitations as a low specificity and the need for human intervention because it is currently not automated [43].

Estimating unmeasurable quantities. Determining reasonable estimates for currently unmeasurable bioprocess variables requires the use of so-called *soft sensors*, consisting of mathematical models constructed for the express purpose of inferring the unmeasured values using available measurements generated by analytical devices (i.e. 'hard' sensors). The approach's premise is that the unmeasured variables are connected to the measured ones according to some known — and 'invertible' — mathematical relationships, so that the unknown quantities can be solved for in terms of the known quantities.

Soft sensors, which can take a variety of forms, have been used successfully in chemical manufacturing for decades, ranging from relatively simple static 'calibration equations' to dynamic state estimators using Kalman filter [44–47]. Not so for bioprocesses. The use of soft sensors for estimating unmeasured states in cell cultures or CQAs remains rare (See, for example, [48,49•]; and for a comprehensive review of Kalman filter applications in microbial processes, see [50].).

Future needs

The lack of high-precision, real-time sensors for productivity and product quality attributes argues for the development of novel sensors capable of measuring these quantities efficiently and at desired frequencies.

Because of the complexity of the attributes in question, it may be necessary to deploy not a single sensor for each attribute, but an array of sensors that will produce intermediate information, requiring further processing to infer the values of the individual attributes. The design of such sensor arrays (what types of sensors; how many of each; how and where to deploy them geographically around the process, etc.) along with the associated inference algorithms, will require innovations in process analytical fundamentals, hardware, and software.

In light of the numerous challenges yet to be overcome in developing robust glycan sensors, in the meantime, it may be useful to consider the design of soft sensors that can estimate glycan distributions using mathematical models.

Control

What and why

From one perspective, controlling bioprocesses is similar to controlling chemical, mechanical, or any other type of process where a control objective is achieved by manipulating appropriate process inputs. Most biopharmaceutical processes are fed-batch processes, not continuous. With no steady state around which to control the process, and with process conditions changing dynamically and nonlinearly from the beginning of the run until the end 10–15 days later, controlling a bioprocess shares some common features with controlling batch or semi-batch chemical processes, and is similarly more challenging than controlling continuous chemical processes.

From another perspective, however, bioprocess control has its own distinctive features and unique challenges — with a defining characteristic being that the actual manufacturing 'unit' in a bioprocess is *not* the bioreactor; it is the cell, a living entity with its own complex operational objectives and its own internal regulatory mechanisms. The implications are enormous. For example, unlike in the more familiar case of chemical process control, the living manufacturing units in a bioprocess are numerous, heterogeneous, and distributed throughout the bioreactor; more importantly, the micro-scale subprocesses within each cellular unit cannot be controlled *directly* with the manipulated variables available to the bioreactor operator at the macro scale. Furthermore, violating important operational constraints can be catastrophic for the living cell.

Effective control of bioprocesses is important for many reasons. The two most important are:

1. To support cell proliferation and productivity, the necessary nutrient levels and stable bioreactor conditions, must be maintained. With cells consuming nutrients, producing metabolites, and with the pH/DO/osmolality of culture media changing constantly, this objective can only be achieved through active and effective automatic process control. Manual control is ineffective and inefficient.
2. Demonstrating a robust process capable of manufacturing products that meet strict quality standards *repeatably* and *predictably*, is a regulatory requirement — necessitating effective process control.

Current status

Current approaches to bioprocess control fall into four broad categories:

1. regulatory control of base bioreactor conditions;
2. regulatory control of nutrient and metabolite concentrations;
3. open-loop, model-based control of product attributes; and
4. closed-loop, model-based control of product attributes.

In the first category, the primary control objective is to maintain pH, DO, temperature, and stirring speed at prespecified setpoints, by manipulating the mass flow rates through the pumps, heater power, and motor power. While such base-level regulatory control is critical to the success of a run, its implementation technology is mature and not unique to bioprocesses; it needs no further discussion.

The objective in the second category is to induce and maintain certain metabolic behavior in cells (e.g. inhibiting lactate production) by maintaining nutrients or metabolites constant at prespecified levels. An increasing body of work on PID control of nutrient and metabolite levels uses Raman probes to measure glucose, lactate, ammonia, CO_2, cell density, and viability, generating readings multiple times per hour (without manual sampling), which are then used for feedback control. In [51], the glucose level in a culture medium was controlled using a proportional controller, while in [52•], the glucose *and* the lactate levels were controlled *simultaneously* using two PI controllers; likewise in [53], where the glucose and amino levels were controlled. In contrast to conventional bolus feeding strategies, these examples show that controlled and continuous supplementation of feed media can reduce fluctuations in nutrient and metabolite concentrations, with positive effects on the final product characteristics.

In the third category, the few recent attempts at model-based control of product attributes are considered 'open-loop' control because the manipulated variables were not adjusted in automatic feedback mode. Instead, the optimal control action sequence was determined *before* running the bioprocess and implemented *verbatim* during process operation without feedback. For example, in [55], to reduce batch-to-batch variability, the final biomass of a recombinant protein was controlled by implementing control action computed *à-priori* using a dynamic bioprocess model.

In [56•], a dynamic glycosylation model was used to optimize the feed rates and nutrient concentrations to increase the galactosylation level of a mAb product. More recently, in [57], multiple glycosylation attributes were optimized simultaneously using a multiscale glycosylation model. These examples of open-loop control represent a first step toward fully automated model-based feedback control of productivity and CQAs in mAb

production. However, the absence of feedback means that any effect of plant/model mismatch or process disturbances cannot be compensated for during process operation.

Applications in the fourth control category are currently few in number. In [54], model predictive control (MPC) was used for maintaining minimal nutrient concentrations during the exponential growth phase of the cell culture as this was deemed to be crucial in reducing the lactate and ammonia concentrations. In [59], MPC was used to control the fraction of high-mannose antibodies produced in a perfusion cell culture, with the MPC algorithm using measurements from daily samples to determine the appropriate amount of mannose supplement to be added into the culture medium. More recently, in [60•], MPC was also used to control lactate concentration of a cell culture in a fed-batch bioreactor. An auto-regressive model was used to represent the process dynamics, with pH and the nutrient feed rate, the exogenous inputs, updated daily based on lactate measurements. While these examples represent commendable achievements in closed-loop, model-based control of mammalian cell culture processes, only a single attribute was controlled in each case.

Future needs

A closed-loop, model-based, multi-attribute bioprocess control system does not currently exist. To meet industry needs, at a minimum, such a system should have the following capabilities

1. real-time monitoring of bioreactor conditions, nutrient levels, and various product attributes, using fast-turn-around analytical methods and/or soft sensors;
2. reliable dynamic predictions of process output variables; and based on these,
3. dynamic, optimal adjustment of process inputs to meet control objectives.

All indications are that MPC, appropriately tailored to the peculiar characteristics of bioprocesses, offers the best option. However, despite its wide adoption across a broad spectrum of other manufacturing applications, MPC is yet to penetrate biomanufacturing broadly. Until the aforementioned modeling and estimation challenges are resolved effectively, the current status of limited MPC applications in biomanufacturing is unlikely to change.

Another future need is the seamless integration of control software with bioreactor hardware. Currently, multiple communication protocols are available for different equipment, posing a major challenge for a seamless integrating with the software on which mathematical models and control algorithms are developed [61]. Open Platform Communications (OPC), a manufacturer-agnostic series of universal communication protocols, offers a potential solution to this problem. It features easy connectivity between devices and continues to receive wide acceptance from both hardware and software vendors. Several recent studies report the use of OPC for data acquisition and control [51,52•].

Conclusions

In manufacturing therapeutic mAbs and other biologics, it is crucial but challenging to achieve desired productivity and product quality precisely, efficiently, and consistently. It is our opinion that a PSE-based approach, comprising three components, process modeling, estimation, and control, can address (and already is addressing) many of the major challenges faced by the biopharmaceutical industry. We argue for a hybrid approach to modeling, which involves employing a judicious combination of mechanistic knowledge and data for model development. Novel sensors capable of measuring the bioprocess quantities of interest, efficiently and at desired frequencies are needed, along with model-based soft sensors to infer whatever unmeasurable quantities remain.

There is much promising progress being made in process modeling, estimation, and control, which, collectively, are building toward the ultimate objective of routine implementation of fully automated systems for optimal bioprocess operations (as is currently the case with the chemical manufacturing industry). There are also formidable challenges yet to overcome.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest

1. *Biopharmaceutical Market — Forecast (2021–2026). Technical Report*. IndustryARC; 2019.
2. Lu R-M, Hwang Y-C, Liu I-J, Lee C-C, Tsai H-Z, Li H-J, Wu H-C: **Development of therapeutic antibodies for the treatment of diseases**. *J Biomed Sci* 2020, **27**:1-30.
3. Xu J, Xu X, Huang C, Angelo J, Oliveira CL, Xu M, Xu X, Temel D, Ding J, Ghose S *et al.*: **Biomanufacturing evolution from conventional to intensified processes for productivity improvement: a case study**. *Mabs, vol 12* 2020:1770669.
4. Torkashvand F, Vaziri B: **Main quality attributes of monoclonal antibodies and effect of cell culture components**. *Iran Biomed J* 2017, **21**:131.
5. Chang MM, Gaidukov L, Jung G, Tseng WA, Scarcelli JJ, Cornell R, Marshall JK, Lyles JL, Sakorafas P, Chu A-HA *et al.*: **Small-molecule control of antibody n-glycosylation in engineered mammalian cells**. *Nat Chem Biol* 2019, **15**:730-736.
6. Kontoravdi C, del Val IJ: **Computational tools for predicting and controlling the glycosylation of biopharmaceuticals**. *Curr Opin Chem Eng* 2018, **22**:89-97.

7. Villiger TK, Scibona E, Stettler M, Broly H, Morbidelli M, Soos M: **Controlling the time evolution of mab n-linked glycosylation. Part ii: model-based predictions**. *Biotechnol Prog* 2016, **32**:1135-1148.

8. Gao J, Gorenflo VM, Scharer JM, Budman HM: **Dynamic metabolic modeling for a mab bioprocess**. *Biotechnol Prog* 2007, **23**:168-181.

9. Baughman AC, Huang X, Sharfstein ST, Martin LL: **On the dynamic modeling of mammalian cell metabolism and mab production**. *Comput Chem Eng* 2010, **34**:210-222.

10. Kontoravdi C, Pistikopoulos EN, Mantalaris A: **Systematic development of predictive mathematical models for animal cell cultures**. *Comput Chem Eng* 2010, **34**:1192-1198.

11. Krambeck FJ, Betenbaugh MJ: **A mathematical model of n-linked glycosylation**. *Biotechnol Bioeng* 2005, **92**:711-728.

12. del Val IJ, Nagy JM, Kontoravdi C: **A dynamic mathematical model for monoclonal antibody n-linked glycosylation and nucleotide sugar donor transport within a maturing golgi apparatus**. *Biotechnol Prog* 2011, **27**:1730-1743.

13. Jedrzejewski PM, Del Val IJ, Constantinou A, Dell A, Haslam SM, Polizzi KM, Kontoravdi C: **Towards controlling the glycoform: a model framework linking extracellular metabolites to antibody glycosylation**. *Int J Mol Sci* 2014, **15**:4492-4522.

14. Radhakrishnan D: *Modeling, Estimation, and Control of Glycosylation in Monoclonal Antibodies Produced in CHO Cells (Ph.D. thesis)*. University of Delaware; 2016.

15. Spahn PN, Hansen AH, Hansen HG, Arnsdorf J, Kildegaard HF, Lewis NE: **A markov chain model for n-linked protein glycosylation-towards a low-parameter tool for model-driven glycoengineering**. *Metab Eng* 2016, **33**:52-66.

16. Hutter S, Villiger TK, Brühlmann D, Stettler M, Broly H, Soos M, Gunawan R: **Glycosylation flux analysis reveals dynamic changes of intracellular glycosylation flux distribution in Chinese hamster ovary fed-batch cultures**. *Metab Eng* 2017, **43**:9-20.

17. Kremkow BG, Lee KH: **Glyco-mapper: a Chinese hamster ovary (cho) genome-specific glycosylation prediction tool**. *Metab Eng* 2018, **47**:134-142.

18. Sha S, Huang Z, Agarabi CD, Lute SC, Brorson KA, Yoon S: **Prediction of n-linked glycoform profiles of monoclonal antibody with extracellular metabolites and two-step intracellular models**. *Processes* 2019, **7**:227.

19. Martínez JA, Bulté DB, Contreras MA, Palomares LA, Ramírez OT: **Dynamic modeling of cho cell metabolism using the hybrid cybernetic approach with a novel elementary mode analysis strategy**. *Front Bioeng Biotechnol* 2020, **8**:279.

20. Yahia BB, Gourevitch B, Malphettes L, Heinzle E: **Segmented linear modeling of cho fed-batch culture and its application to large scale production**. *Biotechnol Bioeng* 2017, **114**:785-797.

21. Zürcher P, Sokolov M, Brühlmann D, Ducommun R, Stettler M, Souquet J, Jordan M, Broly H, Morbidelli M, Butté A: **Cell culture process metabolomics together with multivariate data analysis tools opens new routes for bioprocess development and glycosylation prediction**. *Biotechnol Prog* 2020, **36**:e3012.

22. Powers DN, Trunfio N, Velugula-Yellela SR, Angart P, Faustino A, Agarabi C: **Multivariate data analysis of growth medium trends affecting antibody glycosylation**. *Biotechnol Prog* 2020, **36**: e2903.

23. Kotidis P, Kontoravdi C: **Harnessing the potential of artificial neural networks for predicting protein glycosylation**. *Metab Eng Commun* 2020:e00131.

24. Grainger RK, James DC: **Cho cell line specific prediction and control of recombinant monoclonal antibody n-glycosylation**. *Biotechnol Bioeng* 2013, **110**:2970-2983.

25. St. Amand MM, Hayes J, Radhakrishnan D, Fernandez J, Meyer B, Robinson AS, Ogunnaike BA: **Identifying a robust design space for glycosylation during monoclonal antibody production**. *Biotechnol Prog* 2016, **32**:1149-1162.

26. Yang O, Ierapetritou M: **mab production modeling and design space evaluation including glycosylation process**. *Processes* 2021, **9**:324.

27. Abt V, Barz T, Cruz-Bournazou MN, Herwig C, Kroll P, Möller J, Pörtner R, Schenkendorf R: **Model-based tools for optimal experiments in bioprocess engineering**. *Curr Opin Chem Eng* 2018, **22**:244-252.

28. von Stosch M, Hamelink J-M, Oliveira R: **Hybrid modeling as a QbD/PAT tool in process development: an industrial E. coli case study**. *Bioprocess Biosyst Eng* 2016, **39**:773-784.

29. • Zhang D, Del Rio-Chanona EA, Petsagkourakis P, Wagner J: **Hybrid physics-based and data-driven modeling for bioprocess online simulation and optimization**. *Biotechnol Bioeng* 2019, **116**:2919-2930

A hybrid modeling framework was developed for bioprocess on-line monitoring, prediction, and optimization. High-quality data were generated using a physics-based noise filter and used subsequently to develop a data-driven model. This hybrid framework was shown to be predictive and flexible.

30. • Luo Y, Lovelett RJ, Price JV, Radhakrishnan D, Barnthouse K, Hu P, Schaefer E, Cunningham J, Lee KH, Shivappa RB *et al.*: **Modeling the effect of amino acids and copper on monoclonal antibody productivity and glycosylation: a modular approach**. *Biotechnol J* 2020:2000261

A hybrid modeling technique for efficient, model-based, bioprocess design and control was developed and validated experimentally. Such a technique allows one to update an existing dynamic process model with new supplemental inputs without modifying its structure.

31. Jackson R, Bayrak ES, Wang T, Coufal M, Undey C, Cinar A: **High performance agent-based modeling to simulate mammalian cell culture bioreactor**. *Computer Aided Chemical Engineering, vol 44* 2018:1453-1458.

32. Bayrak ES, Wang T, Jerums M, Coufal M, Goudar C, Cinar A, Undey C: **In silico cell cycle predictor for mammalian cell culture bioreactor using agent-based modeling approach**. *IFAC-PapersOnLine* 2016, **49**:200-205.

33. Soheilypour M, Mofrad MRK: **Agent-based modeling in molecular systems biology**. *BioEssays* 2018:1800020.

34. Ghahramani Z: **Probabilistic machine learning and artificial intelligence**. *Nature* 2015, **521**:452-459.

35. Chakraborty A, Sivaram A, Samavedham L, Venkatasubramanian V: **Mechanism discovery and model identification using genetic feature extraction and statistical testing**. *Comput Chem Eng* 2020, **140**:106900.

36. Rafferty C, O'Mahony J, Rea R, Burgoyne B, Balss KM, Lyngberg O, O'Mahony-Hartnett C, Hill D, Schaefer E: **Raman spectroscopic based chemometric models to support a dynamic capacitance based cell culture feeding strategy**. *Bioprocess Biosyst Eng* 2020, **43**:1415-1429.

37. Capito F, Zimmer A, Skudas R: **Mid-infrared spectroscopy-based analysis of mammalian cell culture parameters**. *Biotechnol Prog* 2015, **31**:578-584.

38. Li M-Y, Ebel B, Paris C, Chauchard F, Guedon E, Marc A: **Real-time monitoring of antibody glycosylation site occupancy by in situ Raman spectroscopy during bioreactor cho cell cultures**. *Biotechnol Prog* 2018, **34**:486-493.

39. Doherty M, Bones J, McLoughlin N, Telford JE, Harmon B, DeFelippis MR, Rudd PM: **An automated robotic platform for rapid profiling oligosaccharide analysis of monoclonal antibodies directly from cell culture**. *Anal Biochem* 2013, **442**:10-18.

40. Dong J, Migliore N, Mehrman SJ, Cunningham J, Lewis MJ, Hu P: **High-throughput, automated protein a purification platform with multiattribute lc-ms analysis for advanced cell culture process monitoring**. *Anal Chem* 2016, **88**:8673-8679.

41. Chi B, Veyssier C, Kasali T, Uddin F, Sellick CA: **At-line high throughput site-specific glycan profiling using targeted mass spectrometry**. *Biotechnol Rep* 2020, **25**:e00424.

42. Zhang L, Luo S, Zhang B: **The use of lectin microarray for assessing glycosylation of therapeutic proteins**. *MAbs, vol 8* 2016:524-535.

43. Guerra A, von Stosch M, Glassey J: **Toward biotherapeutic product real-time quality monitoring**. *Crit Rev Biotechnol* 2019, **39**:289-305.

44. Ogunnaike BA: **On-line modelling and predictive control of an industrial terpolymerization reactor**. *Int J Control* 1994, **59**:711-729.

45. Fortuna L, Rizzo A, Sinatra M, Xibilia MG: **Soft analyzers for a sulfur recovery unit**. *Control Eng Pract* 2003, **11**:1491-1500.

46. Vieira GNA, Olazar M, Freire JT, Freire FB: **Real-time monitoring of milk powder moisture content during drying in a spouted bed dryer using a hybrid neural soft sensor**. *Drying Technol* 2019, **37**:1184-1190.

47. Yongfeng FU, Hongye S, Jian C: **Mimo soft-sensor model of nutrient content for compound fertilizer based on hybrid modeling technique**. *Chin J Chem Eng* 2007, **15**:554-559.

48. Ohadi K, Legge RL, Budman HM: **Development of a soft-sensor based on multi-wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures**. *Biotechnol Bioeng* 2015, **112**:197-208.

49. • Narayanan H, Behle L, Luna MF, Sokolov M, Guillén-Gosálbez G, Morbidelli M, Butté A: **Hybrid-ekf: hybrid model coupled with extended kalman filter for real-time monitoring and control of mammalian cell culture**. *Biotechnol Bioeng* 2020, **117**:2703-2714

A soft sensor for mammalian cell culture processes was developed using a hybrid model and an extended Kalman filter. The authors demonstrated the potential of using soft sensors to measure some process variables of interest accurately and frequently.

50. Yousefi-Darani A, Paquet-Durand O, Hitzmann B: *The Kalman Filter for the Supervision of Cultivation Processes*. 2020.

51. Berry BN, Dobrowsky TM, Timson RC, Kshirsagar R, Ryll T, Wiltberger K: **Quick generation of raman spectroscopy based in-process glucose control to influence biopharmaceutical protein product quality during mammalian cell culture**. *Biotechnol Prog* 2016, **32**:224-234.

52. • Eyster TW, Talwar S, Fernandez J, Foster S, Hayes J, Allen R, Reidinger S, Wan B, Ji X, Aon J *et al.*: **Tuning monoclonal antibody galactosylation using Raman spectroscopy-controlled lactic acid feeding**. *Biotechnol Prog* 2020:e3085

The promise of Raman spectroscopy as a reliable on-line monitoring technology for mAb production processes was demonstrated in this paper. Raman technology enabled control of glucose and lactate in a CHO cell culture using a continuous feed stream of glucose and lactic acid. As a result, compared with what was obtained using a bolus feed, the metabolite concentrations were steadier, the ammonium levels were reduced, and the galactosylation levels were increased.

53. Webster TA, Hadley BC, Dickson M, Busa JK, Jaques C, Mason C: **Feedback control of two supplemental feeds during fed-batch culture on a platform process using inline Raman models for glucose and phenylalanine concentration**. *Bioprocess Biosyst Eng* 2020:1-14.

54. Frahm B, Lane P, Atzert H, Munack A, Hoffmann M, Hass VC, Pörtner R: **Adaptive, model-based control by the open-loop-feedback-optimal (olfo) controller for the effective fed-batch cultivation of hybridoma cells**. *Biotechnol Prog* 2002, **18**:1095-1103.

55. Jenzsch M, Gnoth S, Beck M, Kleinschmidt M, Simutis R, Lübbert A: **Open-loop control of the biomass concentration within the growth phase of recombinant protein production processes**. *J Biotechnol* 2006, **127**:84-94.

56. • Kotidis P, Jedrzejewski P, Sou SN, Sellick C, Polizzi K, del Val IJ, Kontoravdi C: **Model-based optimization of antibody galactosylation in cho cell culture**. *Biotechnol Bioeng* 2019, **116**:1612-1626

An open-loop, model-based control strategy to increase the galactosylation level of a mAb product was developed and validated experimentally. This paper represents a first step toward fully automated model-based control of productivity and CQAs inmAb production.

57. Kappatou CD, Ehsani A, Niedenführ S, Mhamdi A, Schuppert A, Mitsos A: **Quality-targeting dynamic optimization of monoclonal antibody production**. *Comput Chem Eng* 2020, **142**:107004.

58. Amand St, Tran K, Radhakrishnan D, Robinson AS, Ogunnaike BA: **Controllability analysis of protein glycosylation in cho cells**. *PLOS ONE* 2014, **9**:e87973.

59. Zupke C, Brady LJ, Slade PG, Clark P, Caspary RG, Livingston B, Taylor L, Bigham K, Morris AE, Bailey RW: **Real-time product attribute control to manufacture antibodies with defined n-linked glycan levels**. *Biotechnol Prog* 2015, **31**:1433-1441.

60. • Schmitt J, Downey B, Beller J, Russell B, Quach A, Lyon D, Curran M, Mulukutla BC, Chu C: **Forecasting and control of lactate bifurcation in chinese hamster ovary cell culture processes**. *Biotechnol Bioeng* 2019, **116**:2223-2235

A closed-loop, model-based control strategy for maintaining the lactate concentration in a CHO cell culture was developed and validated experimentally. This paper represents a commendable achievement in closed-loop, model-based control of mammalian cell culture processes.

61. González I, Calderón AJ, Figueiredo J, Sousa J: **A literature survey on open platform communications (opc) applied to advanced industrial environments**. *Electronics* 2019, **8**:510.