Author/s:
Yatipanthalawa, BS;Gras, SL

Title:
Predictive models for upstream mammalian cell culture development: a review

Date:
2024-03

Citation:
Yatipanthalawa, B. S. & Gras, S. L. (2024). Predictive models for upstream mammalian cell culture development: a review. Digital Chemical Engineering, 10, https://doi.org/10.1016/j.dche.2023.100137.

Persistent Link:
https://hdl.handle.net/11343/339523

Digital Chemical Engineering 10 (2024) 100137

Contents lists available at ScienceDirect

Digital Chemical Engineering

journal homepage: www.elsevier.com/locate/dche

Review Article

Predictive models for upstream mammalian cell culture development - A review

Bhagya S. Yatipanthalawa [a,b], Sally L. Gras [a,b,*]

[a] Department of Chemical Engineering, The University of Melbourne, Parkville, VIC 3010, Australia
[b] The Bio21 Molecular Science and Biotechnology Institute, The University of Melbourne, Parkville, VIC 3010, Australia

ARTICLE INFO

Keywords:
Cell culture modelling
Mathematical modelling
Mammalian cell culture
Digital bioprocessing
Hybrid models
Biopharmaceutical production

ABSTRACT

The production of therapeutic proteins in mammalian cell culture is an essential unit operation in biopharmaceutical manufacture that can benefit from the predictive insights of effective process models, leading to accelerated process development and improved process control. This review outlines and evaluates current approaches to predictive model development for mammalian cell culture and protein production. Classical mechanistic and data driven approaches are analysed, together with potential challenges in model development and application, including the experimental requirements for parameter estimation. Hybrid models, which may offer greater robustness, are then explored along with hybrid model architecture and the steps involved in model development. Successful examples from other cell fermentation processes are also considered, for application to the development, monitoring and control of mammalian processes.

1. Introduction

The application of digital techniques to therapeutic protein production by mammalian cell lines is a promising area, where there is potential to increase process understanding, accelerate process development and provide greater process control. Advances and insights directed at processes using Chinese hamster ovary (CHO) cells are of particular significance, as these cells are commonly used for therapeutic protein expression by the biopharmaceutical industry, due to their ability to carry out post translational modifications, such as protein glycosylation, protein folding and assembly that are critical to product quality (Kyriakopoulos et al., 2018).

Digital transformation is attractive, as it offers the potential to develop solutions that might not be possible through experimentation alone, whilst also reducing the high costs associated with experimentation. Smart processes are expected to learn from the data generated by bioprocesses and predict optimal states. Further, digital technologies and predictive modelling can be used to improve the quality and reproducibility of products by design, i.e. a quality by design (QbD) approach rather than quality by testing approach (Sommeregger et al., 2017)..

This review identifies and analyses the different approaches used to develop predictive models for cell culture in the current literature, addressing a key challenge identified for digital biotransformation of model development (Smiatek et al., 2020). It focuses on the development of models for mammalian cell culture that describe the upstream steps involved in monoclonal antibody (mAb) production, with a focus on processes using CHO cells. It first examined how cell culture models may be applied, before briefly reviewing the different statistical or data driven approaches to modelling mammalian bioprocesses. Major mechanistic approaches (structured or unstructured, as well as segregated or unsegregated) are then examined in detail. The challenges associated with developing mechanistic models, including experimental requirements and parameter determination are then explored, together with the advantages and disadvantages of each mechanistic approach. The potential for hybrid models, including their different configurations is finally outlined, along with the promise of coupling model development and refinement with a design of experiments framework. Learnings from other cell culture types and the potential application of digital bioprocessing to an industrial context are also considered to give an overview of promising research directions.

2. How cell culture models may be applied for digital transformation

Models of mammalian cell culture may be used to address several key

* Corresponding author.
E-mail address: sgras@unimelb.edu.au (S.L. Gras).

https://doi.org/10.1016/j.dche.2023.100137
Received 25 September 2023; Received in revised form 1 December 2023; Accepted 20 December 2023
Available online 24 December 2023

operational challenges that reduce or limit current or new manufacturing processes as part of a broader effort to achieve digital biotransformation. This includes:

1) Reducing process uncertainty and perturbation to optimise both product quality and yield, including a reduction of fluctuations that impact negatively on cell culture performance (Park et al., 2021, Sommeregger et al., 2017).
2) Optimising the trade-off between productivity, cell density and product quality (Radhakrishnan et al., 2018).
3) Identifying the interdependence of key performance indicators (KPIs) [titre, yield] and critical quality attributes (CQAs) [quality through glycosylation patterns, purity through the presence of by-products] with culture environments (O'Flaherty et al., 2020). This requires the adjustment of key & critical process parameters (KPPs and CPPs), including media composition and bioreactor conditions [pH, temperature, dissolved oxygen (DO)] (Park et al., 2021) to achieve optimal processes.
4) Reducing the time spent on process optimization (Li et al., 2010).

Once developed and validated, such models will help in process understanding, process state estimation, output prediction, as well as in process control (Luo et al., 2021) and may also be used for the development of a "digital twin" that can predict cell growth and productivity under dynamic conditions.

An appropriate model type must first be identified, however, which requires knowledge of the possible types of cell culture models and the advantages and disadvantages of different modelling approaches, which is covered in the next two sections.

3. The classification of different cell culture models

Classification of different cell culture models can be based on the level of process knowledge incorporated or the degree of hybridization (Narayanan et al., 2022), whether the model considers dynamic or static predictions and whether the model is stochastic or deterministic (Ben Yahia et al., 2015). Fig. 1 outlines these possible classifications for modelling cell culture.

Mechanistic models incorporate mechanistic knowledge, i.e., an underlying mechanism, whereas data driven, or statistical approaches, merely depend on the relationship between the model inputs and outputs without mechanistic insight (De Alwis et al., 2007). A hybrid approach combines these two strategies (Von Stosch et al., 2014). All three of these different approaches can, to a certain extent, be divided into either stochastic or deterministic models, as well as being dynamic or static models.

A stochastic model accounts for the uncertainty inherent in the system, incorporating probabilistic based variation within the model input variables, this approach can mainly be seen when considering the cell population in a cell culture (Mantzaris, 2006). Deterministic models, by contrast, are mainly based on experimental observations and do not consider the heterogeneity of the cell population and rather consider the average behaviour of the cell population (Ben Yahia et al., 2015).

Static and dynamic models are differentiated based on whether temporal variation is included within a model (Park et al., 2021). Static models are simpler, focus on a specific instance or end point of the cell culture and typically consist of algebraic equations, whereas dynamic models observe the evolution of the modelled system over time and involve differential algebraic expressions that are computationally more demanding but offer greater process insights (Noll and Henkel, 2020). Hence, the broader differentiation of the three different cell culture modelling approaches can include a stochastic/ deterministic approach and can also consider temporal variation or exclude this level of detail.

In the following sections, we focus first on statistical or data driven models and then mechanistic models, which are the main approaches currently used to represent and understand complex bioprocesses (Farzan et al., 2017), before considering the alternative hybrid approaches outlined in Fig. 1.

3.1. Statistical or data driven models

With the growing availability of bioprocess data, statistical models are increasing in number. These are mainly data driven black box approaches that are based on statistical correlations between the process variables and outputs, which in most cases do not account for an underlying biological mechanism (Sha et al., 2018). Whilst some of these models may be relatively easy to construct and can offer new insights, they can be limited if there is no true causal relationship between the input and output data. Nevertheless, these are useful tools in a data rich environment, when very little information is known about the system and can be used to obtain novel insights or to predict more complex behaviour compared to first principle models. The presence of many process conditions (e.g. temperature, pH) and dynamically varying variables (e.g. substrate, cell and product concentrations and properties) in mammalian cell culture processes and the inherent complexity of the datasets generated by the biopharma industry makes statistical techniques, such as multivariate analysis, more attractive, as these techniques enable the exploration, visualization and interpolation of complex relationships and patterns (Rathore and Singh, 2015).

Multivariate data analysis (MDVA) techniques have been used to model mammalian cell cultures to identify opportunities for process optimization, control targets and key correlations, often using historical data. Principal component analysis (PCA) and Partial least squares regression (PLS) are two of the most common MVDA methods used in

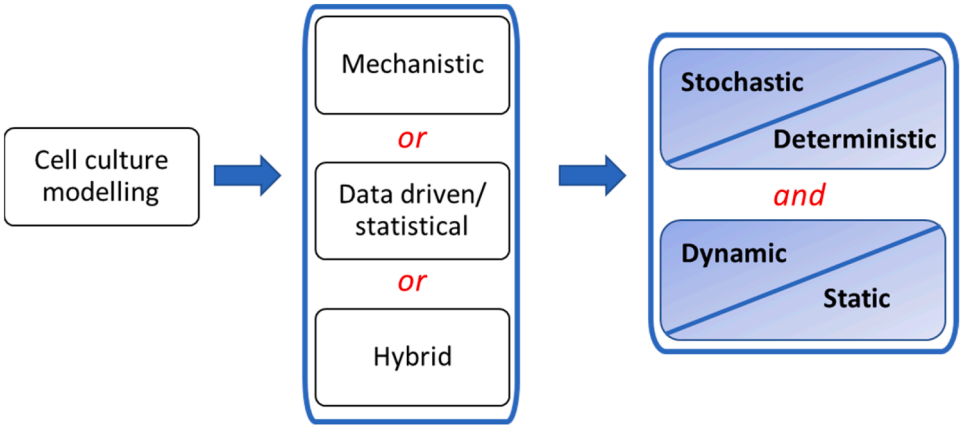


Fig. 1. Classification of the three main different modelling approaches for cell culture and four further types of models for each category.

bioprocess development that rely on dimensionality reduction. This approach is used to create new uncorrelated variables that can represent and explain the variation within the system, whilst retaining important information (Jolliffe and Cadima, 2016). For example, a PCA model with two principal components was used to characterize the impact of media selection on antibody glycosylation and productivity, successfully explaining more than 90% of the variability in the antibody product titre and glycosylation profiles (Fig. 2 (Powers et al., 2020)). Such tools are useful when not much is known about the fundamental mechanism underlying the process and they can be used to effectively capture complex dynamics using models with a relatively simple structure. For instance, in Fig. 2, the optimization of product titre was shown to potentially have deleterious effects on product quality under selected conditions.

Pattern recognition through statistical analyses may also be used as a guide to enhance process robustness and to understand the correlations among complex processes. For example, multivariate analysis of cell culture data has revealed an important correlation between lactate consumption and final product titre for CHO cell cultures (Le et al., 2012). The conditions that induce lactate consumption in high titre runs, however, cannot be completely understood using such statistical techniques.

Data driven models may also incorporate details of the cell culture, such as the presence of different metabolic phases e.g., the lag, exponential and decay phases during a batch or fed-batch culture. For instance, in a successful example as the critical process parameters (i.e. parameters having a higher impact on the weights) observed in a support vector regression model developed to predict the final titre using time series process data, varied when considering five different specific stages of cell culture (Charaniya et al., 2010), indicating that process adjustments may be used to maximise titre by considering cell culture stage. MVDA has also been coupled with existing knowledge on the amino acid stoichiometric balances in CHO cell culture for media development, in order to improve the cell growth and productivity (Salim et al., 2022). Such examples, where process knowledge is coupled with data analysis techniques in a hybrid approach are discussed later in detail.

Powerful data driven approaches, related to machine learning (ML), such as artificial neural networks, have also been used to predict cell growth, titre and product quality, such as the glycoform distribution (Kotidis and Kontoravdi, 2020). Such ML algorithms train and learn from datasets in an adaptive manner that simulates human-like expertise in pattern prediction and decision making (Fisher et al., 2020) and these algorithms are often compared to biological neural networks (Guresen and Kayakutlu, 2011). These are often preferred over the traditional MVDA methods described above, due to their ability to account for the nonlinear interactions between inputs and outputs, leading to improved model robustness.

Many types of different ML algorithms have been used in upstream mammalian cell culture process development and monitoring such as artificial neural networks, tree based regression and gradient boosting techniques. A list of machine learning algorithms and their specific advantages and disadvantages relating to modelling the scale up of bioreactors has been reviewed recently (Alavijeh et al., 2022). An artificial neural network, for example, has been able to accurately predict site-specific glycoform distributions of up to eighteen glycan species, with an average absolute error of 1.1% (Kotidis and Kontoravdi, 2020). This model used experimental or calculated intracellular concentrations of nucleotides and NSDs (nuclear receptor SET domain-containing proteins), extracellular metabolites or gene expression levels for specific

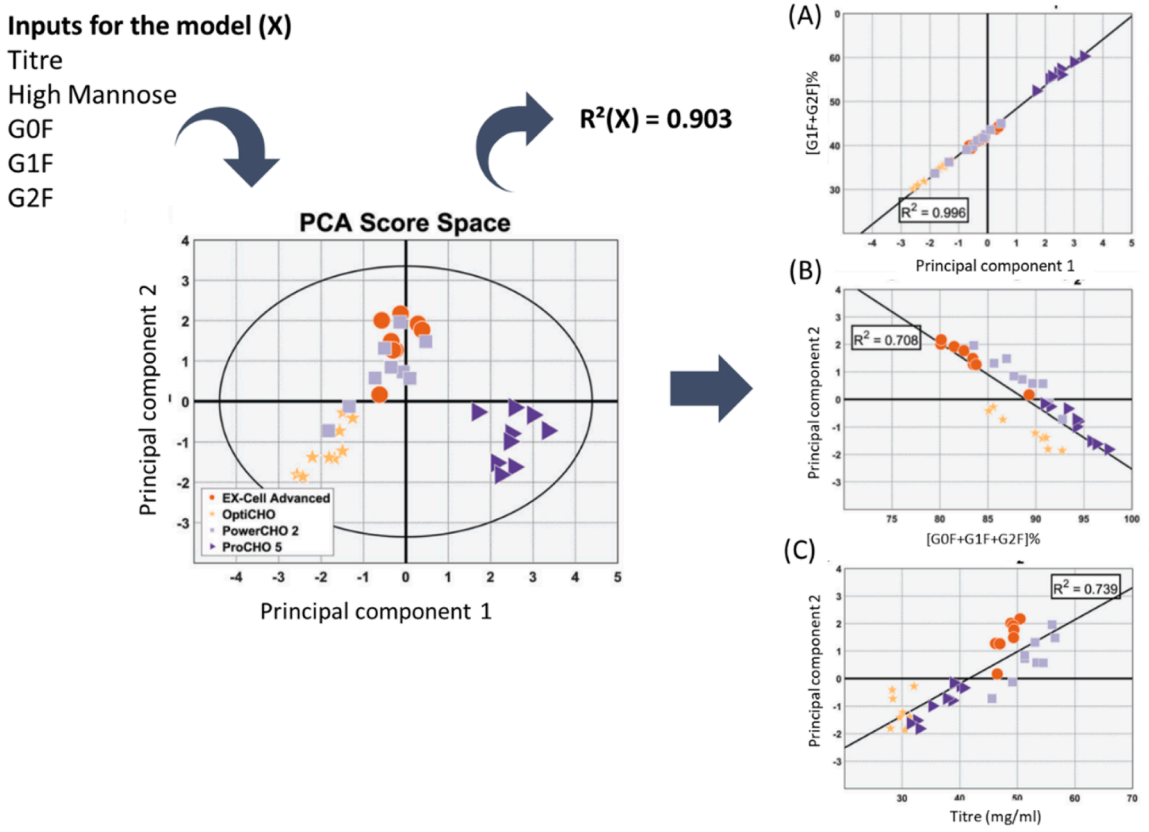


Fig. 2. An overview of the principal component analysis (PCA) model developed by Powers et al. (2020) to characterize the impact of media selection on glycosylation profile and titre. The two principal components cover ~ 90% of the variability in the input data. The first principal component is correlated with terminal galactosylation (A). The second principal component is correlated with the ability to convert high mannose glycoforms into the G0F glycoform (B) and this is inversely correlated to the metabolic phenomena responsible for increased protein production/ titre (C).

glycosylation enzymes, as input variables, or features, to the model, adjusting as needed for different recombinant proteins.

Tree-based regression techniques have been found to outperform neural network and gradient boosting techniques for real time process control and the prediction of product quality in capturing and polishing mAb using chromatography (Nikita et al., 2022), exhibiting optimal performance for smaller process scale datasets with low chances of overfitting, whilst being computationally inexpensive. It is worth noting, however, that the performance of each algorithm may depend on the size, variation and the quality of the specific data set, given that the hyperparameters of the model are tuned and optimised.

The major advantages and disadvantages of the two main data driven techniques discussed above, MVDA and ML, are outlined in Table 1, together with the suitable scenarios for the application of each technique in manufacturing biopharmaceuticals.

Models based on data driven approaches alone are commonly known as black box models, as they are solely based on data driven models without any further incorporation of process knowledge. This leads to a limitation, as black box models are best applied to process conditions similar to those used to develop the model, requiring large data sets at different process conditions in order to improve the model range (Mears et al., 2017). Methods are being developed, however, for obtaining good outcomes from smaller datasets (Sun et al., 2022, Pham et al., 2023). The data needs to be easily accessible, representative and reliable in order to build training datasets and train the model. There can be several challenges associated with the collection of such data for mammalian cell culture processes (Fisher et al., 2020), including:

1. The limitation and variability of data – this can include inconsistencies in data collection; sampling frequency, sensor malfunction, missing data, the presence of outliers that are not representative of the system and measurements that are not at the required granularity. Noting that most measurement frequencies are daily for cell growth. Sampling frequency in mammalian cell culture is further discussed in detail in section 5.1.3
2. The variability in the manufacturing systems – the use of different experimental systems for data collection might lead to unseen variation in the data collected.
3. The selection of manual or automated data collection – manual data collection is time consuming and can introduce human error. Automated data collection increases the volume of data but can introduce challenges with storing and handling large data sets and may still require manual annotation or curation.

Therefore, careful thought needs to be given to the design of experiments used to collect the data for data driven models. Further, the selection of the features that will go into building the data driven model is also important, to ensure the model captures the variability inherent in the process. A key drawback of data driven models is in the inability to extrapolate beyond the input dataset (Fisher et al., 2022), a knowledge of system boundaries is therefore important when developing and using the model for predictions (Fisher et al., 2020, Fisher et al., 2022).

3.2. Mechanistic models

Mechanistic models will often be a combination of fundamental first principle models based on physical processes including mass balances and empirical models used to describe metabolic rates and growth kinetics (Mears et al., 2017).

A fundamental knowledge of cell culture, cellular processes and cell kinetics is central to the development of mechanistic models to predict the behaviour of cell cultures. Compared with data driven or black box models (section 2.1), models developed using mechanistic approaches or knowledge are flexible and can be applied for many purposes (Mears et al., 2017). Their drawback is that the model development requires a significant investment of time, resources and process insight (Mears et al., 2017).

Mechanistic models are based on known relationships and theoretical foundations of a system, based on either physical, chemical or biological knowledge. Coupled ordinary differential kinetic equations make the mathematical framework, describing the time dependant concentration of species and biomass (Smiatek et al., 2020). These can be coupled with flow rates, pH values or other growth considerations, for a more detailed representation of the experimental conditions. Cell processes including metabolism, growth, the cell cycle, cell death and mAb production are all essential interlinked biological processes in mammalian cells (Grilo and Mantalaris, 2019) that may be included in mechanistic models. Such models can be used to study the time dependant concentration profiles of cells, metabolites, products and by-products. Mechanistic models can further be categorised based on the level of physical or biological detail included in the model. The following section focuses on the different mechanistic modelling approaches described in the literature.

4. What are the different mechanistic approaches?

Mechanistic models are primarily categorised (Park et al., 2021) into two categories as outlined in Fig. 3:

1) Structured or unstructured based on the level of intracellular biological detail.
2) Segregated or unsegregated based on the level of heterogeneity of the cell population.

Structured models consider the structure, defined metabolic pathways and cellular compartments (Solle et al., 2017). Unstructured models consider the cell as a whole, whilst excluding intracellular structures and reactions from analysis. A segregated model aims to

Table 1
Comparison of main data driven techniques used along with their advantages, disadvantages and the suitable scenarios where these techniques may be applied.

Data driven techniques	Different approaches	Advantages	Disadvantages	Suitable scenario
Multivariate data analytic techniques (MVDA)	• PCA analysis • PLSR	• Relatively easy to construct. • Exploration, visualization and interpolation of complex relationships. • Relies on dimensionality reduction.	• Only considers linear interactions, non-linear relationships are not considered. • Loss of information during transformation of dimensionality.	• Identifying opportunities for process optimization, control targets and key correlations through pattern recognition. • Predictive modelling.
Machine learning	• ANN • Tree based algorithms. • Gradient based algorithms	• Ability to account for the non-linear interactions between inputs and outputs. • Improved model robustness.	• Depends on the size, variation and the quality of the specific data set, given that the hyperparameters of the model are tuned and optimised. • High training time and computationally demanding. • Selection of suitable/appropriate structure of the algorithm.	• Predictive modelling. • Forecasting.

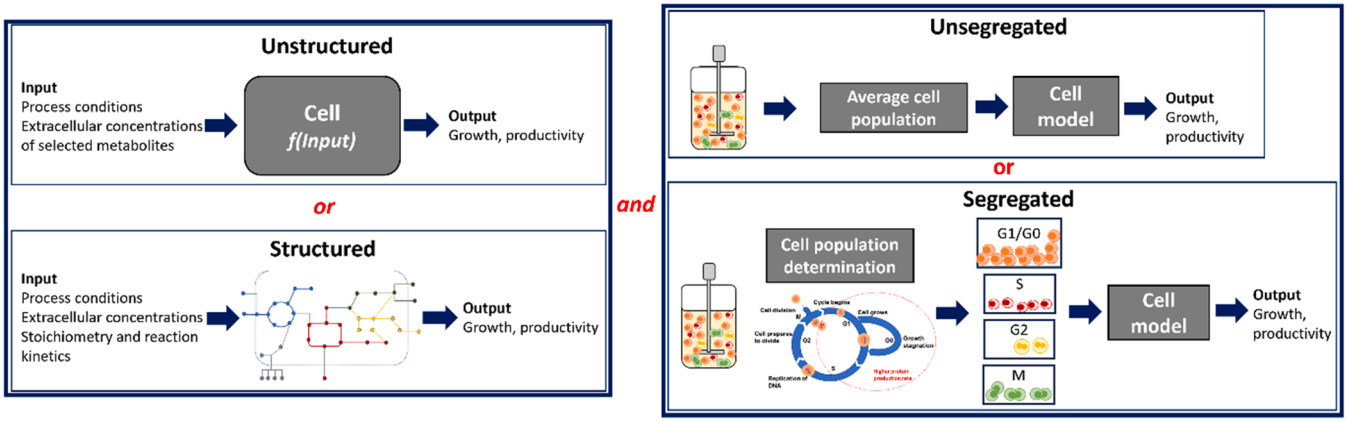


Fig. 3. Classification of mechanistic models based on the level of intracellular biological detail and the heterogeneity of the population considered.

capture the age distribution, or cell cycle, of the cell population (Grilo and Mantalaris, 2019), whereas unsegregated models do not consider the age distribution of cells, rather consider it as a homogenous mixture of cells (Solle et al., 2017).

There have been several attempts in the existing literature to build mechanistic models that consider both the cell structure, i.e. structured models and the inhomogeneity of cell ages within the population, i.e. segregated models, as well as models that consider only one of these two aspects. Other models do not consider any of these aspects i.e. unstructured unsegregated models. The following sections outline the available structured, unstructured, segregated and unsegregated models for mammalian cell culture. For the ease of characterisation and explanation, the models are considered in the following order: structured dynamic models, unstructured dynamic models followed by segregated and unsegregated approaches respectively. It is important to note, however, that these approaches can be combined to explain both the cell structure and the population, forming the categories of structured segregated, structured unsegregated, unstructured segregated and unstructured unsegregated.

4.1. Structured models

Structured models consider the intracellular components and reactions when predicting cell growth or mAb production. These can also be described as single cell models (SCM), meaning they consider the detailed characterization of a single cell (Sidoli et al., 2004). Mammalian cells have a complex internal structure, involving many highly regulated biochemical processes occurring in various locations within the cell (Sidoli et al., 2004). The level of detail considered in structured models, however, may vary. Some models only consider parts of cell physiology, glycolysis, the citric acid cycle (TCA cycle) or both, connecting the extracellular metabolites, such as glucose and amino acids, with intracellular metabolites. Some other models have attempted to go a step further and divide the cell mass to different intracellular metabolic pools (Batt and Kompala, 1989). All these approaches require extensive knowledge of the biochemistry of the mammalian cell and additional experimental measurements to inform the model, including the measurement of the concentrations of intracellular metabolites.

An example of this type, is the CHO genome scale metabolic models, containing a mathematical representation of the metabolic reactions within cells that have been developed with the advancement in multi-omics datasets to describe cell culture behaviour (Hefzi et al., 2016, Selvarasu et al., 2012). Different intracellular metabolic reactions are considered and used to generate a metabolic network model. These genome scale metabolic network models are capable of describing metabolic conversion, from the nutrients in the media to intracellular processes, such as cell proliferation and recombinant protein production (Hefzi et al., 2016). They are therefore identified as a promising platform for *in-silico* monitoring and the prediction of amino acid concentrations required to control product quality and quantity (Yeo et al., 2020). Such genome scale models can also be used to identify cell engineering targets that allow efficient utilisation of resources, although this area is beyond the scope of this review.

Flux balance analysis (FBA) of genome scale metabolic models (GEMs) can also be used to elucidate the physiological behaviours and metabolic states of mammalian cells under various environmental conditions. Such models have been successfully used to identify optimal feed targets, as well as feed media components and to maximise protein production (Huang et al., 2020). For example, a genome scale model of CHO cells was used to understand amino acid metabolism, linking metabolism with IgG production; amino acid supplementation was then used to optimise protein production (Huang et al., 2020). A disadvantage of these detailed mechanistic models, however, is that they will be product, cell line, cell stage and culture condition dependant (Sommeregger et al., 2017, Coulet et al., 2022). For example, the production and secretory pathways may be product specific and differ for the same cell line, requiring reconstruction of the biochemical reactions involved for each product (Gutierrez et al., 2020). While transcriptomic, proteomic and metabolomic data has been integrated with a global genome scale metabolic model for CHO cells to successfully generate models, these are also cell line specific for CHO-K1, CHO-S and DG44 cells (Hefzi et al., 2016).

The advantage of using metabolic flux models for media optimization, is that a range of amino acids and other metabolites can be studied and compared to identify the limiting substrates in the medium. Further, several approaches have been used for industrial application for improving feeding strategies, media costs and sparging strategies (Park et al., 2021, Schinn et al., 2021b). Apart from the inherent difficulty in building such metabolic networks, these models assume an optimal highly efficient metabolism, so can underestimate amino acid consumption (Schinn et al., 2021b). The complexity and the lack of information on cellular processes, including transport, control of enzyme activities and expression or post transcriptional processing of proteins, makes these models hard to develop. There have been several attempts to reduce the size of GEMs in order to improve interpretability and to give physiologically consistent flux distributions (del Val et al., 2021, Ramos et al., 2022). Such a compact genome scale model with only 144 reactions (del Val et al., 2021), has been able to perform comparatively well in predicting the growth rates of cell lines (with only a 2.5% reduction in the predictive capability), whereas a comprehensive genome scale model may have more than 6000 reactions (Hefzi et al., 2016). Existing CHO GEMs might also require modifications and corrections (Yeo et al., 2020), for example a CHO GEM was created (Hefzi et al., 2016) and then subsequently modified (Yeo et al., 2020), by

cleaning the GEM to remove duplicated reactions and metabolites and incorporating enzyme kinetic information.

FBA and metabolic flux analysis (MFA) are based on pseudo steady state mass flux balances that assume the intracellular metabolites do not accumulate in the cell; they are able to provide a detailed picture of the distribution of internal fluxes at a specific condition but are not able to account for the dynamic nature of a cell culture (Martínez et al., 2020). They alone cannot predict the regulation of the fluxes, nor the relationship of the intracellular fluxes with the extracellular environment [20], although data collected at multiple conditions can be used to build a more informative picture. In contrast, kinetic or dynamic metabolic models contain nonlinear relationships, which are more complex than the linear relationships in metabolic flux based models that assume steady state with no accumulation (Kyriakopoulos et al., 2018). For a detailed understanding on the difference between FBA, MFA and kinetic models, the reader is directed to a recent comprehensive review (Sha et al. (2018)).

It is important to note that structured models will preferably be dynamic models (Batt and Kompala, 1989), that combine flux models with dynamic bioreactor models, incorporating external stimuli and their impact on flux (Huang et al., 2020). This is shown in Fig. 4, where the reactions and reaction stoichiometry in a defined metabolic network is considered, a metabolic flux analysis has then been carried out for each reaction relating to all the metabolites. The rate of change in the metabolite is then obtained by linking it to a material balance over the bioreactor. The FBA of a GEM allows intracellular fluxes to be calculated using the optimisation of an objective function, developed based on the final intention/ application of the model. The most commonly used objective function, which aims to maximize biomass production, however, leads to an overestimation of the growth rates (Schinn et al., 2021a). This implies that careful consideration is required in defining the objective function to avoid inaccurate predictions when using FBA combined with GEMs (Park et al., 2010). Future combination of these models with data driven approaches may also provide opportunity to further optimise these models. For example, a tool has already been developed which looks at combining ML/statistical models built using omics datasets providing predictions for new interactions between genes and proteins where there is a physiochemical uncertainty (Erdem and Birtwistle, 2023), allowing the expansion of the capacity of mechanistic models by integrating additional knowledge from big datasets, specifically having applications in the early drug development stages. More relevant to process development, a method combining a principle component analysis (PCA) with a flux balance analysis on a reduced GEM of CHO cells resulting in a *hybridFBA*, which was shown to have potential for the optimization of CHO cell growth through feed optimization (Ramos et al., 2022).

The level of detail in structured models is high and many parameters are involved. Nevertheless, there are several successful examples where structured kinetic models have been used to optimise different upstream processes. For instance, one novel approach, involved integrating a mechanistic metabolic model with subcomponent models for cell growth, signalling regulation and the bioreactor environment (O'Brien et al., 2021). Another kinetic metabolic model was developed that included glycolysis, the pentose phosphate pathway, TCA cycle, respiratory chain, redox state and energetic metabolism that could simulate cell growth kinetics, extracellular glucose, glutamine, lactate and ammonium concentration, as well as the energetic state of the cell (Ghorbaniaghdam et al., 2013). This was used as an *in-silico* model to analyse the effect of butyrate on the behaviour of the CHO cell metabolic network once metabolic regulation was included (Ghorbaniaghdam et al., 2014). A genome scale metabolic network has also been coupled to the dynamics of a chemostat to optimize the media used in a continuous process (Pérez-Fernández et al., 2021). A further model that couples intracellular metabolism during continuous culture to extracellular variables, has been used for media optimization (Fernandez--de-Cossio-Diaz et al., 2017). This specific example considers the growth capacity of the cell and the impact of toxic by product accumulation. Key differences in metabolic function and efficiency have also been observed in CHO cell cultures under batch and fed batch operating conditions, using a dynamic model of CHO cell population dynamics coupled with a kinetic model for cytosolic glucose metabolism (Chen et al., 2014).

These structured models make great platforms for upstream process development, optimisation and control, once parameterised and validated appropriately, they attempt to replicate the internal processes in the cell (Sidoli et al., 2004). Model development, however, is complex and this complexity can also lead to over parameterization, where the model will have more parameters than necessary to accurately describe the underlying data, which can lead to overfitting.

4.2. *Unstructured models*

In unstructured models, cell growth and protein production are matched to extracellular parameters, such as the concentrations of nutrients and metabolites, while the intracellular reactions are not considered (Ben Yahia et al., 2015). Monod type kinetics have been used together with a basic understanding of the production process. A better performing model can be developed when additional information is included on factors affecting:

1) Cell growth and death in batch/fed batch or continuous cultures, and
2) Antibody protein production.

Specific growth rate in an unstructured model can be expressed using different mathematical formula including the Monod equation. An example of this type is an unstructured kinetic model used to predict the specific growth rate of yeast cells and ethanol fermentation, where the least squared error between the model solutions and the experimental data was minimised to ensure fit (Kasbawati et al., 2018) the equations of Monod, Tiessier, Aiba and Ghose and Tyagi type models were compared for model performance (Table 2). An unstructured model, with an Aiba-type structured model component for the specific growth rate of the relevant yeast species, in this case *Saccharomyces cerevisiae*,

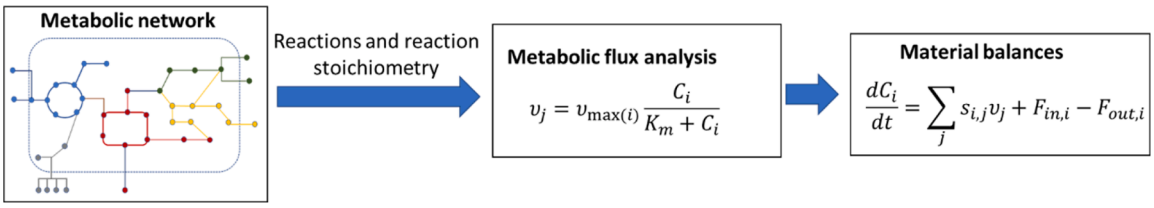


Fig. 4. An example of a structured model combining a genome scale metabolic network model (far left), through which the reactions and reaction stoichiometries are obtained (right). For each reaction in the metabolic network, metabolic flux analysis is carried out to identify the reaction rates, v_j. Considering Michaelis Menten type equations, the reaction rate can be expressed as a function of the Michaelis Menten constant, K_m, maximum reaction rate (v_{max}), and the concentration of the metabolite (C_i). The rate of change of each metabolite in the system can then be obtained through a material balance, where the term $s_{i,j}$ represents the stoichiometric coefficient obtained using the reaction stoichiometry, $F_{in,\,i}$ represents the rate at which metabolite i is fed into the system and $F_{out,\,i}$ represents the rate at which metabolite i is fed out of the system.

Table 2
Mathematical representation of the different models used to express the specific growth rate in yeast cells as explored in Kasbawati et al. (2018), where μ_m represents the maximum specific growth rate, G and E represent the glucose and ethanol concentrations respectively, K_g represents the saturation constant of glucose and E_c represents the ethanol concentration above which cells do not grow.

Monod	Tessier	Aiba	Ghose and Tyagi
$\mu_m \frac{G}{K_g + G}$	$\mu_m\left(1 - \exp\left(-\frac{G}{K_g}\right)\right)$	$\mu_m \frac{G}{K_g + G}\exp(-K_p E)$	$\mu_m\left(1 - \frac{E}{E_c}\right)$

could best describe the batch ethanol fermentation data (Kasbawati et al., 2018).

Unstructured models of mammalian cell cultures, including CHO cell lines (Craven et al., 2013) and hybridoma cell lines (Amribt et al., 2013), have mostly used Monod type kinetics to explain cell growth rather than the alternatives in Table 2. Instead of using complex metabolic reactions and stoichiometric relationships, the unstructured approach assumes cell growth is primarily dependant on a few growth limiting substrates and growth inhibitory products. Glucose, glutamine and glutamate are most commonly used as the growth limiting substrates in mammalian cell cultures, as illustrated schematically in Fig. 5, whereas lactate and ammonia are products of metabolism, which are often considered as growth inhibitors (Amribt et al., 2013, Gadgil, 2015, Kiparissides et al., 2015). The selection of the best combination of substrates and inhibitors for a model will be cell line specific; the factors chosen should also be easily determinable, for the model to be more useful as a predictive tool with minimal inputs.

An understanding of the details of the system is important to define the limiting substrates and waste products that inhibit cell growth. For instance, an unstructured model was developed to predict the cell growth and mAb production, considering glucose and glutamine as substrates (Amribt et al., 2013). Prior to selecting glucose and glutamine as the growth limiting substrates, as well as lactate and ammonia as growth limiting by-products, a metabolic pathway representing the central metabolism of hybridoma cells, composed of the glycolysis pathway, pentose phosphate pathway, glutaminolysis pathway and the TCA cycle was considered. This case also considered overflow metabolism, involving the incomplete oxidation of an abundantly used energy source (glucose and glutamine in this case) under aerobic conditions, where organic end products, such as lactate and ammonia, are excreted (Amribt et al., 2013). In developing their model, three metabolic states were used to depict metabolism in hybridoma cells, namely as:

1. Respiratory metabolism – where glucose and glutamine are completely consumed for respiration, resulting in a low substrate uptake rate.
2. Critical metabolism – where cells are at their maximum specific growth rate.
3. Overflow metabolism – where excess glucose and glutamine result in the production of associated metabolites.

The specific glucose rates were then defined considering the respiratory and overflow metabolism. This provides a good example of the development of simple unstructured models using expert knowledge.

Different logistic equations have also been used apart from Monod kinetics to predict cell growth and productivity. The constants in these logistic equations, however, have been described as having no meaning compared with Monod constants (Kargi, 2009). Nevertheless, in one example, the Verhulst logistic model for cell growth and productivity in CHO cell cultures was shown to be mechanistic and the model constants found to have physical meaning and biological significance, influencing both cell growth and productivity directly (Shirsat et al., 2015a).

Lysine has on occasion been identified as a growth limiting substrate for myeloma cell growth, which is different to the commonly used growth limiting substrates glucose and glutamine (Liu et al., 2008). In this instance, the Monod model was unable to predict the growth lag phase well, so a metabolic regulator model was combined with the Monod model, yielding good predictions of the specific cell growth rates for myeloma cells, even during the lag growth phase (Liu et al., 2008). Table 3 outlines a summary of the different combinations of growth limiting substrates and growth inhibiting by-products involved in predicting the cell growth using unstructured models in different mammalian cell lines.

Predictions of cell growth and mAb production titre from unstructured models can also be used to predict product quality attributes, such as glycosylation, as both the glycosylation pattern and glycoform distribution can impact mAb quality. For example, a modelling platform that quantifies the impact of feeding precursors that lead to increased glycosylation, such as galactose and uridine, was developed to assess the impact of these precursors on cellular growth, metabolism, antibody productivity and glycoform distribution (Kotidis et al., 2019). The team used the specific growth rate and specific rate of mAb production as inputs to a NSD metabolic model, quantifying the NSD concentration, which then fed to a N-linked glycosylation model to estimate the glycoform distribution.

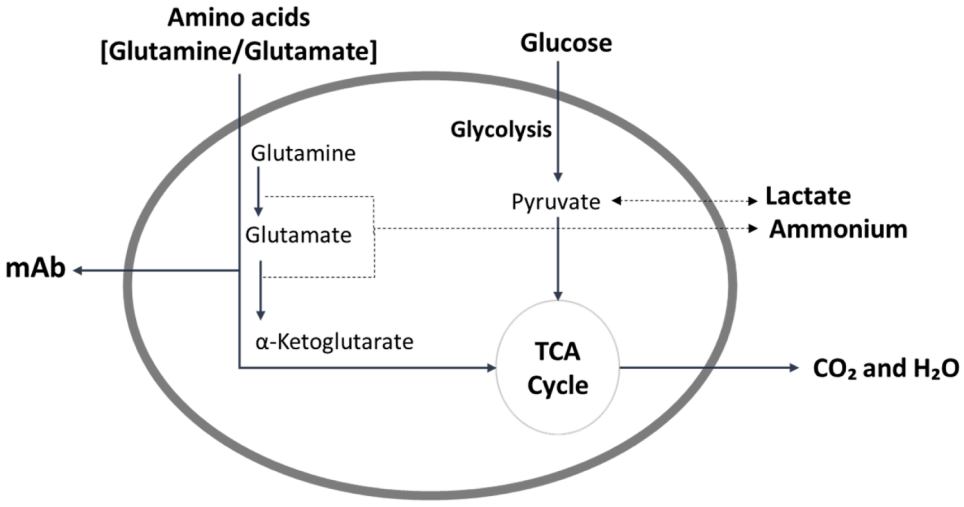


Fig. 5. Schematic of major pathways in CHO cells linked to cell growth and recombinant protein production adapted from Pereira et al. (2018). These pathways can be used for development of unstructured models.

Table 3
Different substrates and inhibitors selected for modelling different mammalian cell lines.

Cell line	Substrates	Inhibitors	References
GS-NS0	Glucose, Glutamate	Ammonia, Lactate	(Kiparissides et al., 2015), (Wang et al., 2022)
	Glucose, Glutamate, Aspartate, Asparagine, Arginine	-	(Papathanasiou et al., 2017)
Hybridoma	Glucose, Glutamine	-	(Amribt et al., 2013)
	Glucose, Glutamine	Ammonia, Lactate	(Fernandes et al., 2015), (Kontoravdi et al., 2010)
	Glutamine	Ammonia, Lactate	(Bree et al., 1988)
	Glutamine, Serum	Ammonia	(Glacken et al., 1989)
CHO	Glucose, Glutamine	Lactate, Ammonia	(Xing et al., 2010), (Shirsat et al., 2015a), (Gadgil, 2015), (Craven et al., 2013)
	Glutamine, Asparagine, and Glucose	Glucose, Lactate, and total Ammonia (protonated and unprotonated)	(Karra et al., 2010)
	Glucose, Glutamine	Ammonia	(Möller et al., 2019)
	Glucose, Glutamate, Glutamine	Lactate, Ammonia	(Xu et al., 2019)
	Glucose, Asparagine	Ammonia, Lactate, Uridine	(Kotidis et al., 2019)
	Glucose	-	(López-Meza et al., 2016)
AGE1.HN. AAT	Glucose, Glutamine	-	(Rath et al., 2014)
Myeloma	Glucose, Glutamine, Lysine	-	(Liu et al., 2008)
CHO, HEK 293	Glucose and essential amino acids (Glutamine, Aspartate, Arginine, Phenylalanine, Serine, Histidine, Valine, Lysine, Threonine, Isoleucine, Leucine)	-	(Kontoravdi et al., 2007)

Prior knowledge of the specific cell line, metabolism and production process are critically important for the development of even the simplest unstructured unsegregated kinetic model for good prediction of cell growth and product formation, as can be demonstrated by the selection of different amino acids and metabolites as substrates and inhibitors that affect the cell growth and productivity. Selection can also depend on the objective of the modelling problem. Therefore, the end purpose of the model needs to be considered early in model design, whether the end use is process optimisation, process characterisation, online monitoring or process control. For instance, glucose and asparagine were considered as growth limiting substrates, while lactate, ammonia and uridine were considered as growth inhibitors for CHO cell cultures, where the objective was to quantify the impact of glycosylation precursor feeding (specifically that of galactose and uridine) on cell growth, metabolism and productivity (Kotidis et al., 2019). In contrast, a good prediction of CHO cell growth was obtained by considering glucose as the only growth limiting substrate, for concentrations over a threshold substrate concentration (0.6 g/l found for the specific CHO cell line; (López-Meza et al., 2016)). The main objective of this second example, however, was to simply characterise the growth kinetics, substrate consumption and product formation in recombinant CHO cell cultures using simple Monod kinetics.

Unstructured models can also be referred to as macroscopic models, as they only look at the macroscopic behaviour of the cells, rather than looking at intracellular detail. This type of model allows integration of biochemical reaction networks and their stoichiometry into models. Although less accurate, these models are easier to set up and to apply. Even though unstructured models give reasonable agreement with experimental and computational data, incomplete knowledge of metabolite reactions and oversimplified considerations in pseudo first order Monod reaction kinetics have been identified as potential shortcomings.

4.3. Segregated vs unsegregated

Segregated models aim to explore the population inhomogeneities of mammalian cell culture that can affect productivity and growth kinetics. This includes the effect of the progression of the cell through the cell cycle, which can affect the dynamics of the cultivation process. For example, cells at different stages of the cell cycle can have different protein expression levels, leading to high and low producing cells (Möller et al., 2020). A cell cycle is the series of steps occurring within a cell between one cell division to the next. As cell cultures are not typically synchronised, they contain a mixture of cells at different stages of the cell cycle. The different stages of the cell cycle, shown in Fig. 6 below, are commonly identified as four steps.

1. G1 phase – cell growth.
2. S phase – cell replication of the genome.
3. G2 Phase – cell growth and preparation for cell division.
4. M phase – includes two steps: mitosis where chromosomes divide between the two daughter cells and cytokinesis, where the cytoplasm of the cell physically divides in two.

Apart from these actively dividing cells, a cell population can consist of cells that have temporarily stopped dividing and reached a state of quiescence, called the G0 phase. In response to external stimuli, some cells may reactivate early response genes, while others may undergo apoptosis or programmed cell death (Karra et al., 2010). The antibody synthesis rate in mammalian cell cultures is observed to be a maximum in the G1/G0 phase and early S phases (Karra et al., 2010, Ramirez and Mutharasan, 1990), making these phases important to production, although all phases are potentially important for modelling.

The fraction of the cell population in different stages of the cell cycle can be identified using fluorescent nucleic acid dyes (Karra et al., 2010), by determining the cell size as an indirect measurement (Ramirez and Mutharasan, 1990) or fractionating the cells by centrifugation (Möller et al., 2018), followed by quantification. For instance, the cell population belonging to the different cell cycle phases was determined as a function of batch time for CHO cell cultures grown in spinner flasks using propidium iodide and RNase followed by flow cytometry (Karra et al., 2010). The fraction of G1/G0 cells increased exponentially from ~0.5 to ~0.8 during the first few hours and then stabilised, while the cell fraction in both S and G2/M cell cycles decreased and stabilised at ~0.1 over the same period. Segregated kinetic models consider and account for these differences in the cell cycle during the development of a cell growth model. These models are also described as population balance models (PBM) in the literature (Sidoli et al., 2004), as they attempt to understand the balance of subpopulations based on cellular state. In contrast to these segregated models, unsegregated models consider the cell population to be homogenous, which is a simplification given the known variation in cell culture, although this assumption still allows effective model development for some situations.

There have been several examples where population balance models introduce multiple sub populations in a mammalian cell culture to address the heterogeneity of the system. These are outlined in Table 4 below.

The development of these segregated models requires additional experiments to identify the different subpopulations in a cell culture and their variation as a function of culture time or as a function of other

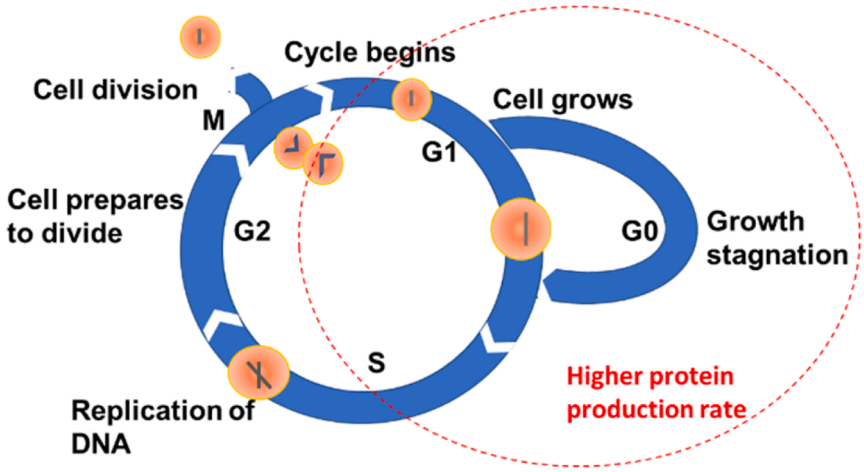


Fig. 6. Schematic representing the various stages in a CHO cell that are considered by segregated kinetic models. Adapted from Karra et al. (2010).

process variables. This area of research is still being developed and it is expected that more accurate technologies will allow the differentiation of the different cell cycle stages and the quantification of their specific growth and protein production rates (Naderi et al., 2011), leading to the development of more reliable segregated models. Currently, fluorescent nucleic dyes can be used to identify the different cell populations, measured using flow cytometry (Meshram et al., 2012) or fluorescence microscopy (Naderi et al., 2011). The significant noise in these measurements can lead to discrepancies in cell state measurement using these techniques, which limits current accuracy.

5. Challenges associated with developing mechanistic models

Whether it is a highly complex structured segregated model or a simple unstructured unsegregated model, there are several challenges associated with developing mechanistic models. Even the simplest mechanistic modelling approach, involving an unstructured unsegregated model, requires good background knowledge and an understanding of the cell culture system which is to be modelled.

Irrespective of the complexity of the mechanistic model, there will always be parameters that need to be estimated or determined based on experimental data. As the number of parameters that need to be determined increases, the complexity of the problem increases. The following sections discuss these aspects in turn.

5.1. Experimental requirements for parameter estimation

To estimate parameters in predictive models, prior experiments are required. The number of experiments required can depend on the number of variables considered in the model, which can be addressed in a design of experiments approach. A further factor is the frequency of measurements, these aspects are considered in the following sections.

5.1.1. Choice of variables

Depending on the complexity of the model, the experimental requirement may vary from measuring a few extracellular metabolite concentrations, such as glucose, glutamine, lactate and ammonia during cell culture (López-Meza et al., 2016) to measuring intermediates of several metabolic pathways, such as citrate, succinate, fumarate, as well as possible genomic, proteomic and transcriptomic analyses (Ghorbaniaghdam et al., 2014). If the heterogeneity of the cell culture is to be considered, additional experiments are required to determine cell age, typically using flow cytometry, to identify caspase activation using a caspase assay (Meshram et al., 2012), giving an indication of the cell status, as previously described above in section 4.3.

The measurement of extracellular metabolites, such as glucose, lactate, glutamate and ammonia is comparatively easier than other metabolites. Spectroscopic assays (Meshram et al., 2012), HPLC techniques (López-Meza et al., 2016) or integrated biochemistry analysers (Selvarasu et al., 2012) can allow rapid analysis of samples collected manually or automatically during cell culture.

If the model considers the intracellular metabolism, intracellular metabolite concentrations, such as organic acids, nucleotide concentrations, amino acid concentrations, DNA and RNA compositions often need to be measured, which requires cell rupture and intracellular metabolite extraction (Ghorbaniaghdam et al., 2013, Nolan and Lee, 2011). This typically involves treating the sample with cold methanol in the presence of sand (Ghorbaniaghdam et al., 2014) or treating with ice cold 50% acetonitrile in milliQ water (Chen et al., 2014) to enable intracellular metabolite extraction. Once extracted, these metabolites can be analysed using high end analytical techniques, such as mass spectrometry. Whilst mass spectrometry is well developed as an analytical platform, this analysis and the subsequent analysis of data require an investment of resources and time. Nevertheless, the advancement in analytical techniques including nuclear magnetic resonance (NMR) (Le Guennec et al., 2012), liquid chromatography (LC), gas chromatography (GC) (Zhang et al., 2013a) and mass spectrometry (MS) (Balcerczyk et al., 2020) has led to novel, accurate and faster methods for measuring metabolite concentration (Shirsat et al., 2015b).

It is important to consider at the outset whether such detailed analyses are required or whether fewer extracellular metabolite and physical factors would suffice for model development.

5.1.2. Design of experiments – number of experiments required

Another consideration for experimentation is the number of experiments that need to be performed, to build a sufficiently accurate model. A design of experiments (DoE) technique can be used to plan and minimise the number of experiments required (Retamal et al., 2018, Bandara et al., 2009). These techniques offer a systematic method to evaluate multiple variables simultaneously. DOE is most often used in process optimisation, within the quality by design approach. This approach identifies the process variables and statistically determines the experimental space with user defined boundaries, which is then experimentally evaluated with respect to the targeted outcomes. Different experimental designs have been used for mammalian cell culture process design and optimisation including full factorial, fractional factorial, box Behnken, central composite design (Zhang et al., 2013b), Taguchi methods (Aghili and Zarkesh-Esfahani, 2018) and even sequential use of two different experimental design methods (Puente-Massaguer et al., 2019). The experimental outcomes are further used to predict the parameters of an empirical response surface model (RSM), which can

Table 4
The application of population balance models in the literature.

Population balance model used	Factors considered	What is explained	Reference
Three stage description of the cell cycle along with a dead state is used and implemented as a time discrete and stochastic cell ensemble model - incorporates G1, S and G2/M phases.	The model focuses on volumetric growth, DNA replication rate and the main limiting substrates (glucose and glutamine) and the main by-product (lactate) for the examined cell cultures.	Cell cycle adaptation and application of a cell cycle dependant population balance ensemble model to describe and understand synchronized bioreactor cultivations in two model mammalian cell lines AGE1 and CHO-K1 for batch and fed batch culture.	(Jandt et al., 2015)
The cell population is divided into three categories: - Viable cycling cells (including S, G2 and M phase cells); - Viable arrested cells (G1/G0 cells); and - Dead cells.	Uses an unstructured model with cell population data to describe cell growth, metabolism of cellular macromolecules and production of mAb	Production of cellular macromolecules and monoclonal antibodies, metabolism of glucose and glutamine with the production of lactate and ammonia, as well as the profiles of cell growth in batch and fed-batch culture.	(Jang and Barford, 2000)
One-dimensional age-based PBM formulated with three flow-cytometry identifiable cell cycle phases; G1/G0, S and G2/M.	The PBM is combined with an unstructured average cell intrinsic state model, which considers the effect of environmental factors (temperature, pH, dissolved oxygen, shear caused by agitation and aeration).	The model explains cell growth, nutrient uptake and metabolite and protein formation in CHO cell batch culture.	(Karra et al., 2010)
Discriminates between viable, apoptotic and dead cells, viable cells are then further segregated based on their cell cycle phase in G0/G1, S and G2/M.	Couples a cyclin and DNA based PBM of cell cycle, an energy based kinetic metabolic model and a gene expression-based apoptosis model.	Predicts cell growth and population heterogeneity (capturing cell cycle and apoptosis), cellular metabolism and mAb production in GS-NS0 batch cultures.	(Grilo and Mantalaris, 2019)
Cell cycle model which considers two states, A and B. State A contains the initial, variable part of the G1 phase with both cycling and arrested cells. State B contains cells that have spent varying lengths of time in state A and includes the S, G2 and M phases.	Cell death rate, which is explained as a function of the number of cells that have arrested within the G1 phase.	Provides a cell cycle model describing a steady state specific death rate, growth rate and cell viability in continuous suspensions of hybridoma cultures.	(Linardos et al., 1992)
Considers the fraction of cells in G1, S, G2 and M phases in a cell culture.	Uses a simple Monod type equation to explain cell growth, considering a cell population balance. mAb productivity is then considered as a function of the fraction of arrested cells in the cell culture.	Applies the cell culture model to a continuous suspension culture to predict the mAb production in hybridoma cell cultures.	(Suzuki and Ollis, 1989)

explain the interactions of the process variables. The systematic planning of experiments through DoE reduces experimental burden, whilst also providing a method to describe the interactions between the process variables and the responses (Möller et al., 2019). Nevertheless, DoE design and execution still requires expert knowledge to determine the set points and boundaries for input factors.

The DOE technique can also be integrated with or replaced by various digital tools to reduce the number of experiments required in bioprocess development for monoclonal antibody production using CHO cell cultures. Recently, the concept of model assisted design of experiments for knowledge-based bioprocess development has been proposed (Möller et al., 2019), to decrease the number of experiments required in mammalian cell culture process development. Here, the user defined boundary values are used to obtain an optimal DoE design, for which the responses are first predicted using a mathematical process model. The predictions are then analysed and screened using prior knowledge, leading to a reduction in the boundary values and therefore experimental space. DOE can also be combined with hybrid models for process characterization, as demonstrated for non-mammalian systems, like *E. Coli* (Bayer et al., 2020c).

Intensified design of experiments, or iDOE, is a more recent approach. This method further reduces the number of experiments required compared to characterisation using a full factorial design space characterization (Nold et al., 2021, Bayer et al., 2021b). This concept is based on introducing intraexperimental process variation (i.e. varying the process mid-way through experimentation), maximizing the information output and enhancing process understanding (Baker et al., 2015). iDoE has been applied to mammalian cell culture by performing sequential setpoint changes during the growth phase for two selected input parameters, temperature and dissolved oxygen. The process performance related outputs (viable cell density (VCD), total cell density (TCD), viability, glucose and lactate concentrations) were then measured (Nold et al., 2021). The data could then be divided into three distinct stages relating to the change in process parameter setpoints. Ordinary least squares models (OLS) were fitted for each output parameter as a function of temperature, dissolved oxygen (DO) and time, for each of the stages used, allowing the effects on process performance to be assessed. A good predictive ability was shown following validation using the data from a classical DoE performed in parallel, demonstrating the potential of iDoE, as a cost-efficient option for mammalian bioprocess development.

A simple genetic algorithm has also been used as an alternative method of experimental design for mammalian cell culture in place of a traditional DoE approach (Brinc and Belič, 2019), potentially addressing the limitations of DoE, including the high dimensionality of the parameter space arising from the complexity of mammalian cell culture processes. This study was able to successfully optimize a set of 14 process parameters within a dataset of 132 experiments per cell line to produce a monoclonal antibody and an Fc-fusion protein using two different cell lines in a fed batch culture, demonstrating the promise of

this approach. The number of experiments required using a traditional DoE will increase drastically if it was to use in optimising 14 process parameters. For example, if a full factorial design was to be used, for optimising 14 process parameters with at least 3 factors considered for the range of each process parameter, a total of 14^3 experiments totalling to 2744 experiments will be required.

5.1.3. Frequency of measurements

The culture of mammalian cells is time dependant due to the cell cycle; processes run in fed batch or semi-continuous mode also vary with time. This dynamic nature requires the experiments in mammalian culture, whether it be for process development or control, to be accurately planned. Mechanistic models typically use ordinary differential equations to explain dynamic variations in cell culture. Parameter estimation in such systems mostly depends on using data fitting by numerical approximation of an initial value problem, such as the Runge Kutta method (Ramsay et al., 2007). Hence obtaining sufficient data points that can capture the variations occurring over the time course, allows suitable estimations for the parameter values to be obtained, as discussed in the next section. Before designing experiments, it is therefore important to have an understanding on the biological dynamics or the temporal characteristics of the system (Kusena et al., 2021a).

Increasing the sampling frequency can have dual impacts on temporal data, as it can increase both:

1) Noise in the experimental data, due to frequent sampling and associated measurement errors.
2) Accuracy of parameter estimation, when using nonlinear ordinary differential equations to explain the variations in cell culture.

Hence, it is important to find the optimal sampling frequency to have sufficient data for parameter estimation, whilst minimising noise (Harrison and Baker, 2018). Smoothing techniques, such as cubic spline have previously been used to reduce the noise in experimental data (Laursen et al., 2007), however, this method should be applied with great care or avoided to ensure any inherent variation in the cell culture is not lost. A sufficient data collection frequency should be identified based on the insights gained through experiments, allowing relevant changes in the cell culture dynamics to be captured.

The metabolism of mammalian cell cultures is typically slower compared to microbial cultures (Reyes et al., 2022). Hence, for mammalian cell culture systems, including CHO cell lines, the dynamics of cell culture and protein production has mostly been sufficiently captured in data points collected at 24 h intervals compared to yeast and other microbes, where data is collected in hourly intervals (Ram et al., 2019). For building a predictive model, however, the required predictive range and accuracy required of the model should also be taken into consideration when identifying the ideal sampling frequency. For example, a higher frequency may be required for real-time process monitoring and control, such as glucose feeding in fed batch reactors (Nickel et al., 2017). One successful example previously used data sampled in 12 h intervals to build a predictive control model that is able to predict the performance of a cell culture up to 12 h ahead (Aehle et al., 2012). This enables the bolus feed to the bioreactor to be adjusted every 12 h; in some circumstances higher frequencies may also be desired.

Calculating the specific rates in cell cultures using time series experimental data is another commonly used strategy applied when building mechanistic, as well as hybrid models. The temporal resolution of data collection has previously been shown to play a critical role in the specific rate calculations for dynamic fed batch experiments using *E.Coli* cultures (Wechselberger et al., 2013). A rule of thumb method for calculating the SNR (signal to noise ratio) has been proposed, which helps to define the optimal sampling interval, as tested with an *E.Coli* system (Wechselberger et al., 2013). A similar approach can help to define a suitable sampling frequency for dynamic fed batch cultures of mammalian cells. Hence, while frequent sampling is preferred for real time monitoring and process control, an optimal sampling frequency will aim to obtain a minimum number of samples that are able to support the construction of an accurate model, covering the time course variation of the considered variables.

5.2. Parameter estimation

Parameter estimation is an essential, yet challenging step during the construction of kinetic models, irrespective of whether parameters are determined experimentally or numerically. Certain parameters, such as the maximum specific growth rate of a cell line, can be estimated using experimentation. With a design of experiments (DoE) approach, the number of experiments required for determining such parameters can be minimised, as discussed in the previous section 5.1.2 (Bandara et al., 2009). Parameters that cannot be determined experimentally, such as the Monod inhibition constant for lactate and ammonia or the glucose maintenance coefficient, will need to be estimated using numerical methods, where an objective function is optimised and subsequently the distance between experimental data and predicted values is minimised. The sum of squared errors is often minimised to enable optimization in this way (Ben Yahia et al., 2015). Where appropriate, weighting is also used in order to obtain a range of values and to represent the data accurately (Retamal et al., 2018, Endrenyi and Tang, 1980). The Nelder – Mead simplex optimization algorithm is also commonly used to minimise the least squares error criterion (Richelle et al., 2022). Alternatively, a maximum likelihood optimization formulation can be used to estimate the model parameters (Kotidis et al., 2019). In this approach, the probability that the mathematical model predicts the experimental measurements is maximised by assigning values to the physical and variance model parameters. The choice of the optimization algorithm depends on the optimization problem, number of variables, constraints, the model, as well as the availability of the different optimization methods in the selected software (Kontoravdi et al., 2007).

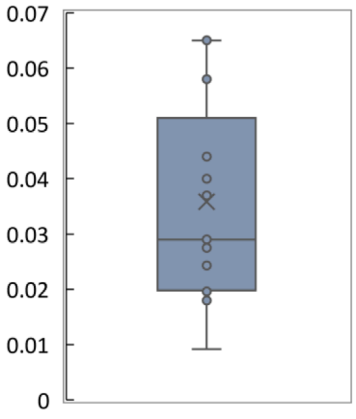
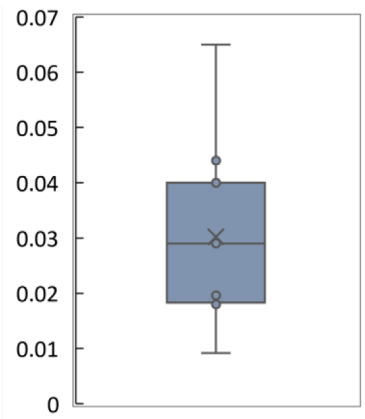
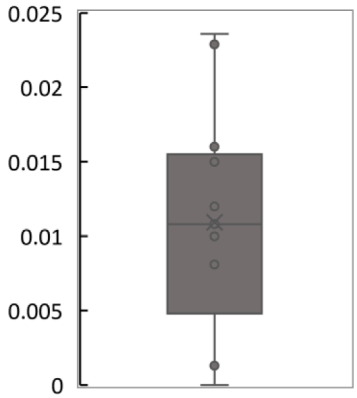
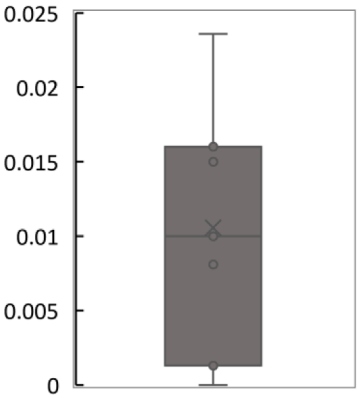
As previously mentioned, the number of model parameters increases with increasing model complexity, leading to difficulties in parameter estimation. The accuracy of the estimated parameters, however, can be improved using prior knowledge (Ben Yahia et al., 2015). An initial guess, as well as upper and lower bounds need to be provided to minimize the least squares error and the literature can guide estimation. An accurate initial guess will help in improving the predictive ability of the models developed, as careful selection can help to converge to a global optimum, instead of being stuck in a locally optimal value. The differences in the cell lines, process parameters and even the units need to be carefully considered, however, during this process.

The range of model parameters obtained in the existing literature is outlined in Table 5, including the maximum specific growth rate and maximum specific death rate, two of the most common parameters used in building unstructured unsegregated models with Monod kinetics. As can be seen, the range varies widely and the initial estimates need to be chosen carefully for the parameter estimation step. It is also interesting to note that the data available for CHO cells include more data with a lower maximum specific growth rate and lower maximum specific death rate than the data collected for a mix of mammalian cell types.

A toolbox has been developed to study and rapidly estimate the specific rates for mammalian cell culture in batch and fed batch cultures using a logistic approach (Goudar, 2012). This can guide and assist bioprocess development activities and enable robust estimation of metabolic fluxes for mammalian cell cultures. In a more recent study, it has been proposed that using cubic spline estimations and obtaining the first derivative of the equations is a simple, yet powerful, way to estimate growth dependent parameters, such as the specific growth rate and specific substrate uptake rate (Bayer et al., 2020a).

An alternative step by step approach has been presented for estimating the parameters in a more complex model (Saraiva et al., 2015). This method starts with a very simple model to obtain parameter estimates and then gradually increases model complexity, using the

Table 5
Range of experimental and computationally obtained values for the maximum specific growth rate and maximum specific death rate in mammalian cells. The general range is shown on the left and for CHO cell lines separately on the right. The boxes depict the interquartile range (range between the first quartile and the third quartile) for values obtained in the literature, while the whiskers represent the range of the data. The median value is represented by the horizontal line inside each box.

Parameter	Mixed mammalian cells	CHO cells only
Maximum specific growth rate (h^{-1})		
Maximum specific death rate (h^{-1})		

parameters estimated in the previous version as the initial estimates for the new model (Saraiva et al., 2015). As discussed previously, the selection of the initial values for the parameters will affect the quality of the predictions. It is therefore recommended that the initial guesses for a system use the existing literature (as illustrated in Table 5). In instances where such reports are scarce, this second type of model could be a more relevant approach to obtain a sufficiently accurate initial guess.

Parameter estimation is highly sensitive to both measurement errors and outliers (Endrenyi and Tang, 1980). In non-linear systems, the major problems associated with parameter estimation are the potential existence of local minima and the practice of over fitting. Once the parameters are identified, the model needs to be assessed for predictive power and for robustness against perturbations. For model validation, the same data set, a different data set, or a combination of datasets has been used, with the latter approaches providing greater insights.

Sensitivity analysis is another important aspect, used to assess model quality. It can provide important information on the importance of parameters in the determination of the model output and possible impact of variability in the model inputs on the output. Not all parameters will have the same impact on the model output. Those that have no affect or have minimal impact on the model output can be identified by performing a sensitivity analysis, thereby reducing the parameter space that needs to be experimentally determined (Kontoravdi et al., 2005). For instance a 'Sobol' global sensitivity analysis can be used to find the parameters that have low sensitivity indices and exhibit strong interactions with one another (Kontoravdi et al., 2010). The probability of a parameter being accurately estimated also increases if it has a strong influence on the model output (Retamal et al., 2018).

Several approaches have been taken for sensitivity analysis. A local sensitivity analysis was performed on each model parameter by changing the value over a range of -85 to +300% while holding the other parameters constant, to determine sensitivity (Ghorbaniaghdam et al., 2013). A derivative based global sensitivity analysis was also able to reduce the number of parameters that need to be experimentally determined from 18 to 9 (Kiparissides et al., 2015). In a more complex study, a sensitivity analysis identified and reduced the parameter space in a dynamic metabolic model of CHO cells (Robitaille et al., 2015). A Morris screening method was used to identify the ranking of the global sensitivity of 139 parameters, returning the distribution of the elementary effect in the space and the most sensitive parameters identified. In this way, a global sensitivity analysis can be used as a tool to quantify the importance of model parameters and their interactions in respect to the model output, leading to optimal experiment design for the subsequent design of experiments required to obtain datasets required for parameter

estimation (Kontoravdi et al., 2005).

Non identifiability is another issue identified in modelling biological systems, which arises due to the highly nonlinear and dynamic nature of the cell culture system used to produce mAb, along with the presence of severely limited measurement of data, which is also noisy (Baker et al., 2015). Non-identifiability relates to whether a unique estimation can be obtained for an unknown parameter within the constraints of the mathematical model. One solution proposed is a unified framework that couples parameter estimation and identifiability analysis (Baker et al., 2015).

Another important aspect that needs to be considered when estimating the parameters is their confidence interval, which determines how much the parameter value is expected to vary from the value estimated. Parameter uncertainty, leading to inaccurate predictions, is an often seen problem in developing kinetic models with nonlinear rate equations (Moreno et al., 2014). Hence estimating the confidence intervals and the correlations between the parameters following parameter estimation is important. The fisher information matrix (FIM) is one method that can be used to obtain a covariance matrix, giving an estimation of the variation, or the uncertainty, of the estimated parameters (Retamal et al., 2018).

Other studies have taken a different approach when investigating the nonlinear dynamic modelling and parameter estimation for a mammalian cell culture process for mAb production (Selişteanu et al., 2015). These include the use of a particle swarm optimization (PSO) based algorithm, which uses time varying acceleration coefficients developed to solve nonconvex optimization problems, giving fast convergence and very good performance. The proposed model, along with the parameter estimation, provided good predictions for cell culture, glutamine, glucose and mAb products. The model also considered the macroscopic reactions, which were linked to mass balances over the reactor.

Variable process parameters are a further complication for parameter estimation, as the parameters estimated are dependent on conditions such as the temperature, pH, seeding density (Kusena et al., 2021b) and osmolality (Alhuthali et al., 2021). While parameters are often assumed to be constant, they can vary slightly during cell culture, leading to errors in model predictions. Modelling approaches that can incorporate these factors into parameter estimation are therefore needed for greater accuracy.

6. Overall comparison of the mechanistic modelling approaches

All four of the mechanistic modelling approaches examined have advantages and disadvantages, which are outlined in Table 6. Despite structured segregated models being the most accurate and closest representation of a cell culture in theory, when it comes to practical implementation, these models impose the most practical challenges. The increased experimental burden, limitations in current analytical techniques, required knowledge of metabolic pathways and the stiff and complex equations can lead to erroneous predictions. For instance, a structured segregated model was not capable of predicting protein production, change in oxygen concentration and osmolality (Bayrak et al., 2015). In this case, a multi agent approach was used, where the CHO cells were considered as individuals/agents following a rule base, which governed their behaviour, while a flux balance model was also applied, where embedded agents predicted quantitative changes in metabolites and nutrients.

Unstructured unsegregated models are the least complex to build among the different mechanistic approaches. These simple models often provide a good starting point for relatively new processes, where data are limited. Such simple models can be preferred in industrial practice, as they are easier to use in process characterization, online monitoring, process control and process optimization. They have been effectively demonstrated for different bioreactor designs (Slininger et al., 2014) and different feeding strategies (López-Meza et al., 2016), which otherwise can require extensive and costly experimentation. Similar models are also commonly used in microbial systems (Zeng, 1995), microalgae and yeast cultivation. While unstructured unsegregated models do not have the ability to capture the effects of the growth condition on cellular composition (Farzan and Ierapetritou, 2018), they can be combined with other models to increase model performance. Some examples include aspects such as protein quality, measured through N linked glycosylation patterns (Villiger et al., 2016) or cellular metabolism (Yahia et al., 2021). For instance, an existing dynamic mathematical model was combined with an unstructured metabolic model to then predict the N linked glycosylation patterns in a fed batch bioreactor (Villiger et al., 2016). This model has been used to quantify the interconnected influences of media components and cell culture conditions.

When developing a predictive model, the major considerations are ease, accuracy and the ability to extrapolate (Smiatek et al., 2020). Achieving the appropriate balance between the ease and accuracy of the predictive model, however, is important. Therefore, unstructured unsegregated models are most used in the literature, as this combination gives the advantage of a predictive model with the ability to extrapolate, coupled with a reasonable, if simplified, understanding of the cell culture system.

7. Towards digital bioprocessing, hybrid models

As discussed in previous sections, models developed using mechanistic approaches are flexible, compared with data driven black box models, with wide potential for application. In contrast, the application of black box models is limited to the conditions used to develop the model and large datasets are required from different process conditions

Table 6
Comparison of the different approaches to mechanistic modelling and their advantages and disadvantages.

	Description	Advantages	Disadvantages
Unstructured unsegregated	• Assumes all cells to be in the same physiological state. Growth and behaviour are affected by extracellular variables only. • Suited to the early stages of process development when limited data is available	• Lower complexity • Faster to develop, easier to handle • Quantitative	• Assumptions and simplifications required. • Limited ability to describe different growth patterns and transitions between growth phases. • Unable to predict product quality (e.g:N glycosylation).
Unstructured segregated	• Considers the cell population to be heterogenous. • Distinguishes between different morphological biomass types or stages.	• Distinguishes between different growth and production behaviour	• Complex with restricted generality • Requires extensive experimental effort
Structured unsegregated	• Considers intracellular processes of mammalian cell cultures and considers the population to be homogenous. • Divides the biomass into compartments with different functions.	• Allows for a sufficient description of the dynamics of cell culture. • Ability to predict protein N glycosylation.	• Requires extensive experimental effort
Structured segregated	• Considers the intracellular processes, whilst treating the population as heterogenous.	• Offers the most realistic representation of a cell	• Heavily parametrised. • Requires extensive effort for both model development and the experimental effort to parameterize the model

but limited mechanistic understanding is required (Mears et al., 2017). While there are advantages and disadvantages associated with each approach, hybrid models developed with the appropriate combination of the two approaches will potentially yield more capable and robust models better able to predict cell culture performance. Hybrid models have gained interest in the fields of chemical synthesis in the prediction of reactions (Venkatasubramanian and Mann, 2022, Bui et al., 2021), as well as in the prediction of cell growth and product formation in biological systems, including mammalian cells (Luna et al., 2021, Thompson and Kramer, 1994). Hybrid models have the advantage of combining fundamental knowledge with empirical data driven methods (Bui et al., 2021). The fundamental component in a hybrid model provides the foundation for insights and extrapolation, whereas the empirical or data driven section accounts for the missing complexity, improving model accuracy (Thompson and Kramer, 1994). These will now be discussed in detail.

7.1 Hybrid models

The concept of hybrid models has existed for a few decades. The earliest reports of a hybrid model for process modelling that combined prior knowledge with data driven approaches, such as artificial neural networks, were seen in the early 1990s (Psichogios and Ungar, 1992, Thompson and Kramer, 1994). Since then, there have been several developments and changes in the way these models are developed and applied to develop and optimise processes. With the advancement in data analytics and availability of data, hybrid models combining mechanistic knowledge and historical process data are increasingly being developed for use as digital tools or in the development of digital twins describing cell culture systems.

Hybrid models consist of two parts: a data driven or black box model and a mechanistic or white box model. These parts may be combined following different architectures, such as a serial or parallel arrangement or a combination of these approaches (Graefe et al., 1999), as shown in Fig. 7. Specifically, there are two possible arrangements in serial and one possible arrangement in parallel, alternatively these approaches may be combined. The term hybrid model is typically used where two distinct mechanistic and data driven models are combined. It should be noted, however, that the hybrid models referenced in this review do not include models that use mechanistic models to generate the data required to train a data driven model (Zhang et al., 2020), as such a model has a strong resemblance to a data driven model (Solle et al., 2017). While these types of hybrid models can be used to explore a larger parametric space to predict the performance of the studied system, with reduced computational power or time, they don't have the advantage of having mechanistic components. For instance, a neural network trained on a limited number of simulations of a mechanistic model has successfully been used to explore a much larger parametric space predicting pattern formation and stochastic gene expression (Wang et al., 2019), overcoming the larger computational time required by complex mechanistic models. Here, models combining data driven and mechanistic model components, along with experimental data, form the focus.

Compared with purely mechanistic models, hybrid models can be more useful, as they offer the flexibility of estimating parameters based on culture conditions (von Stosch et al., 2016). Where the training data are subject to large variations, the use of prior knowledge in hybrid models allows better predictions. Including prior knowledge within the mechanistic component can also reduce the amount of data required to train neural networks, improving the ability of the model to both predict and extrapolate (Thompson and Kramer, 1994).

Hybrid model architecture can be determined based on the research question (Thompson and Kramer, 1994, Solle et al., 2017), with details outlined in a recent review (Solle et al., 2017). Among these different approaches, a serial arrangement is most used in biological systems, where a data driven, method is used to estimate kinetic rates, while a basic material balance is used to incorporate process knowledge (Fig. 7a). Each arrangement can be applied in a different way, to reduce error or to find unknown parameters, as discussed below.

A serial structure, like Fig. 7a, has been used to predict the cell growth and product formation. For instance, a first principle mass balance within a bioreactor was combined in series with an adjustable mixture of neural networks and mechanistic representations of the cell population, to model the growth of baker's yeast (Oliveira, 2004). In such instances, the non-parametric black box component is used to obtain parameters that more realistically represent the culture conditions, including the temperature, pH and feed conditions (von Stosch et al., 2016). The mechanistic component of these models can include information at different levels of detail, as described in Fig. 3, although most models predicting cell growth and product formation have used unstructured unsegregated models for the mechanistic component to date.

Serial structures having a structure like Fig. 7b can be used in instances where a mechanistic/ white box model describes the inner mechanism of the system, while the black box model describes the underlying phenomena that cannot be described in a mechanistic way. For example, a hybrid model has been used to predict product quality, as determined by the glycan distribution profiles of a selected mAb protein in a fed batch CHO cell culture, where the NSD fluxes of the cell are determined by a structured mechanistic model and then used as an input to a neural network (Antonakoudis et al., 2021).

Conversely, hybrid models with a parallel approach (Fig. 7c) are often used to improve the accuracy of the mechanistic model. For instance, a hybrid model was used to improve the accuracy of prediction for intracellular signalling pathways in biological systems, where an artificial neural network was used in parallel with a first principle model to correct predictions (Lee et al., 2020). Hence, if data is available, such hybrid models can be used to improve the predictive capability of structured mechanistic models that might describe genome scale metabolic pathways or metabolic flux analysis.

Successful examples of such hybrid models have not only been used for process development but also for real time process monitoring,

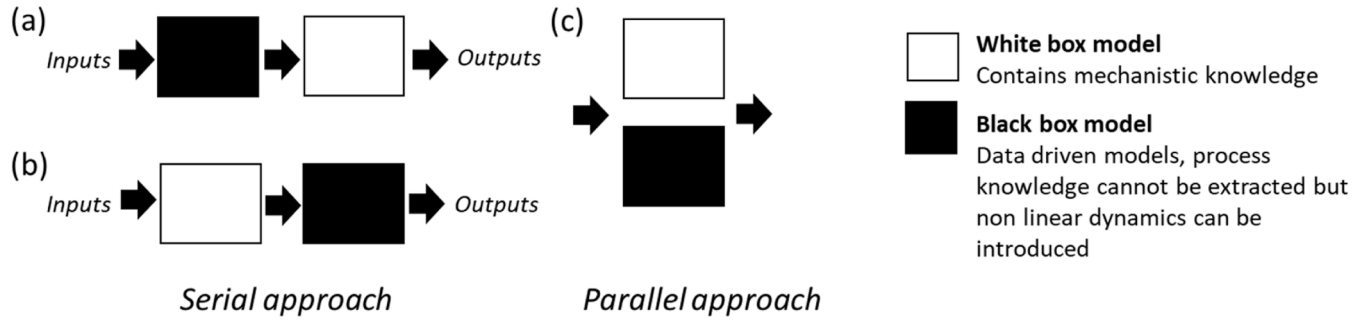


Fig. 7. Schematic representation of the possible architectures for hybrid modelling, where white and black models are combined.

illustrating their potential for application in QbD/PAT tools (Cabaneros Lopez et al., 2021). These examples include the prediction of continuous PHA production by *Pseudomonas putilda* (Luna et al., 2021), the monitoring of cellulose to ethanol conversion in real time for *S. cervisiae* cultures (Cabaneros Lopez et al., 2021), the real time monitoring of growth in mammalian cell culture (Narayanan et al., 2020) and the prediction of sodium gluconate fermentation by *Aspergilus niger* (Dong et al., 2014). Hybrid models have also been used to predict product quality, by predicting mAb galactosylation (Kotidis et al., 2021), suggesting that the drawbacks of a simple unstructured unsegregated model can be overcome using a hybrid approach. Further to this, the ability to account for the varied culture environment conditions, including pH and process temperature, can be an added advantage of a hybrid approach (von Stosch et al., 2016).

The increasing move towards hybrid modelling in industry is also an indication of the promising nature of these models compared with other purely mechanistic, or data driven approaches. Several companies have developed software offerings based on hybrid principles. Approaches vary from the genome scale models of Yokogawa Insilico Biotechnology GmbH to simpler mechanistic models (unstructured unsegregated) (Table 7). In addition to those outlined in Table 7, companies such as DataHow and Siemens PSE also provide platforms for the construction of hybrid models.

There are examples of the data driven/ non parametric component in hybrid models being combined in series (Luna et al., 2021), in parallel (Cabaneros Lopez et al., 2021) or even as a combination of both approaches (Thompson and Kramer, 1994), as shown in Fig. 7. The combined approach using a serial and a second parallel component has been used to predict cell growth, product formation and substrate consumption in a fed batch penicillin fermentation process using penicillin cells, presented in detail in Fig. 8 below (Thompson and Kramer, 1994). This combined approach can suit bioprocesses, as the data are often noisy and subject to high variation. The default model (see Fig. 8) takes in the inputs, including the three state variables for biomass (X), substrate (S), penicillin (P) at time t, as well as the exogenous variables, substrate concentration in the feed S_f, dilution rate (D) and Δt, and produces parameter values, which are then fed into the output model that generates the outputs of biomass (X), substrate (S) and product, penicillin in this case (P) at time $(t + \Delta t)$. The neural network used in this study, a radial basis function network (RBFN), is trained to capture the unknown functional relationships between the inputs and outputs. In this specific case, the cell culture specific parameters used in the default model will need to be known as approximations in the first instance. In scenarios where the specific rate correlations are not completely known as *a priori* equations, the default model in Fig. 8 can be removed, using a serial approach feeding the parameter values from the RBFN or similar neural network to the output model. This has been the approach taken in most of the studies performed to date on biological systems, including mammalian systems (Graefe et al., 1999, Luna et al., 2021). The structure of a hybrid model is therefore amenable to evolutionary model synthesis.

Table 7
Select commercial solutions focussing on hybrid modelling of bioprocesses.

Company	Application	Explanation of Hybrid approach
Yokogawa Insilico Biotechnology GmbH	Provides a process development platform to predict time series growth, titer and metabolite concentrations along with product quality.	The genome scale model, combined with a process model, calculates intracellular fluxes. The machine learning model determines the dynamics between extracellular concentrations, intracellular fluxes and process conditions (temperature and pH) (Nargund et al., 2019).
Novasign	Provides a hybrid model building toolbox.	Uses an ANN to predict parameters using process inputs, the parameters are then fed into a parametric model to obtain the required time series model predictions (product titre/ cell growth). (Bayer et al., 2021a)
Sartorius - Cell insights by Umetrics studio	Provides an in-silico platform to optimize cell culture performance.	Uses a mechanistic model based on Monod kinetics (although accounting for the substrate, inhibitors and a quadratic factor accounting for physical process conditions, such as the temperature and pH) for predicting cell growth along with a hybrid approach for predicting the product titre (using a partial least square method to predict specific productivity, which is then used in a differential equation to obtain the product titre).

Even though the use of hybrid models is appealing, the process of developing such models can be complex and involves several challenges, which are considered here in turn.

First, the appropriate approach needs to be selected, considering whether the data driven and mechanistic components should be combined in serial, parallel or a combination of these approaches. As shown by the examples above, a parallel approach uses a data driven model to alleviate the prediction errors in the mechanistic model, whereas in a serial approach, the data driven model is used to represent the unknown terms or functions in the first principles model, such as the specific rates. The advantage of a serial approach is that it can be relatively simple, using the specific rate term predicted using a data driven approach to account for the varying conditions in the culture environment.

The prediction of the specific rate terms to be used in a serial hybrid approach, however, present a second challenge. These rate terms are not experimentally measured during CHO cell culture, therefore cannot be directly used to train the data driven component. A few approaches have been identified to overcome this challenge (Von Stosch et al., 2014):

1. A direct approach – rate estimation using experimentally measured state values for the concentrations of metabolites, biomass and product (Schubert et al., 1994, Laursen et al., 2007).
2. An incremental approach – which decomposes weighted least squares error between the values for predicted and measured outputs into four smaller problems, which are resolved sequentially (Kahrs and Marquardt, 2008).
3. An indirect approach – uses a sensitivity approach, where error back propagation is performed using sensitivity equations minimising the difference between the values for the concentrations determined experimentally and their predicted outputs (Oliveira, 2004), using gradient based minimization algorithms (von Stosch et al., 2016).
4. A stochastic adaptive moment estimation (ADAM) with semi direct sensitivities approach – a deep learning technique using gradient based adaptive estimates at each iteration (Pinto et al., 2022). This approach is like the sensitivity approach but uses a different method to calculate the gradients, where the sensitivities of state variables in relation to network outputs are calculated, reducing the number of ODEs that must be calculated to determine the sensitivities.

The use of the latter two methods (3 and 4) will be limited to models use ANN for the ML component of the black box model, due to the use of sensitivity approach. In order to obtain nominal values for specific rates, one study fitted experimental data to a mechanistic model first using the complete dataset and then later updated the parameters for each experiment, by refitting them to the mechanistic model equations (Luna

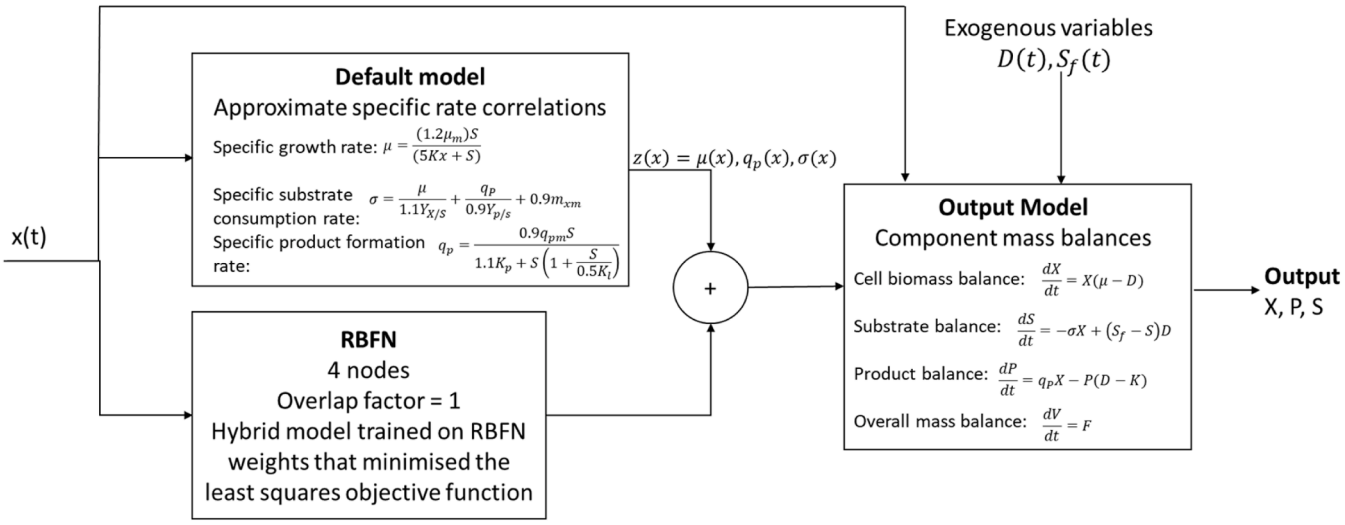


Fig. 8. A schematic representation of a hybrid model developed for penicillin production in a fed batch reactor adapted from Thompson and Kramer (1994)

et al., 2021). Later the substrate concentrations and dilution rates were used to train the neural networks with outputs of specific growth rate, carbon uptake and PHA production rates. Care is needed, as rate estimation using experimental data can lead to distorted estimates if the data used are noisy and if the process has fast dynamics (Schubert et al., 1994). In contrast, the sensitivity approach and other similar approaches can overcome issues with noisy signals but require high intensity computational input.

A further challenge is the appropriate use of data to ensure that maximum information is extracted without overfitting or misinterpretation. Different approaches using artificial neural networks (Narayanan et al., 2019), partial least squares (Carvalho et al., 2022), fuzzy systems (Roubos et al., 1999) and radial basis function networks have all been explored for biochemical systems, including mammalian cell culture systems. The choice of the data driven approach/black box model, choice of hyperparameters when using ML algorithms (von Stosch et al., 2016), inputs to the black box model (Laursen et al., 2007) and the method used to train the model can all contribute to differences in performance. When using the serial approach to obtain the estimates for unknown parameter/rate terms and employing data driven approaches, such as neural networks, a single neural network can be trained to estimate multiple rate terms (Narayanan et al., 2019, Luna et al., 2021) or separate neural networks can be used for each rate term (Laursen et al., 2007, Dong et al., 2014).

The level of process knowledge incorporated into hybrid models can also impact on predictive performance (Narayanan et al., 2022). The addition of more process knowledge can introduce a trade-off between the addition of more model parameters and the increasing need for experimental data to train a robust model (Narayanan et al., 2022). The extent of process knowledge incorporated should therefore be considered based on the target application and availability of data.

Once a hybrid model is built and calibrated for a particular bioprocess, it will need to be recalibrated if it is to be used to predict cell culture performance in a separate setup. Hybrid models have also been applied to transfer knowledge across scales, from 300 ml flask cell cultures to 15 L scale reactors (Bayer et al., 2021b) and across cell lines (Hutter et al., 2021). For instance, a hybrid model has been developed and trained using 300 ml shake flask experimental data on a two-dimensional DoE containing three levels for the cultivation temperature and the glucose concentration of the feed. This model was then used to predict the VCD and product titer in 15 L cultures with the same DoE settings (that is the temperature and glucose feed concentrations) with only minimal recalibration to capture the bioreactor-scale-specific behaviour (Bayer et al., 2021b), although this approach has not yet been demonstrated at larger scale.

Faster bioprocess development strategies may be achieved by coupling of iDoE to hybrid modelling, as iDoE can be used to reduce the experimental burden required to generate the data required to train a newly built hybrid model (Bayer et al., 2020b). The efficacy of an iDoE design space will depend on the ability of the cell culture to respond to changes in the process conditions and careful consideration of the iDoE design space is needed, as described above (section 5.1). Specifically, the conditions should be selected to allow sufficient residence time for the cells to adapt to process conditions (Bayer et al., 2020b). iDoE cannot be used in instances where the combination of certain design factors may lead to irreversible damage to the cells or product quality (Pappenreiter et al., 2023). For example, a recent study examined the impact of the change of various process parameters on the cell growth and productivity, as well as the time taken for the cell line to respond to variations and reach equilibrium; these included responses to temperature shift, pH shift, changes in osmolality and the change in media composition, allowing factors that may irreversibly change the cell metabolism following exposure to be identified (Pappenreiter et al., 2023). Interestingly, shifting the cells from conditioned media to blends of conditioned and unconditioned media resulted in irreversible changes, while mild parameter shifts of pH, osmolarity and increased glucose concentration resulted in reversible changes.

Other factors that need to be considered when using an iDoE design space to train a hybrid model to predict the time course variations in the cell culture include whether:

1. The measurement frequency is sufficient to capture changes in cellular metabolism,
2. Any changes to the specific rates due to the changes in the intra-experimental setpoint variation can be safely captured apart from the inherent dynamic variation of the cell culture,
3. The dataset generated using an iDOE design space may unintentionally introduce bias to the training data,
4. The experimental space of the iDoE, as it is reduced, can capture the full range of interactions within the biological system. For example, if the cell population is not uniform over the culture period and setpoints are changed halfway through the culture, will sufficient information be obtained.

iDOE have been successfully applied to both *E.coli* (Bayer et al., 2020b) and CHO cell lines (Bayer et al., 2021b) but further research is required to assess the applicability of such approaches for transfer across different cell lines and different scales in upstream process development.

The development of open-source platforms may also accelerate progress with hybrid models (Merkelbach et al., 2022). For example, one open-source platform allowed users to capture information in a hybrid model featuring a mechanistic model and a neural network using a JSON file; this approach has shown potential for being flexibly applied to predict drug concentration over time based on physiological information from a given patient in a clinical setting (Merkelbach et al., 2022). Use of such an open-source platform requires the user to have some background knowledge on model development, as well as coding languages but it presents a flexible alternative to the commercial applications outlined in Table 7.

8. Conclusions and future perspectives

Cell culture models capable of predicting cell culture processes are a necessary tool with the current move towards Biopharma 4.0 and the QbD frameworks. Dynamic mechanistic models can provide a good representation of a cell culture, however, parameter estimation can be challenging. Although structured segregated models give the most ideal representation of a cell compared to the other three possible model types, unstructured unsegregated models are often preferred for model development, due to their relative simplicity and their ability to provide reasonably accurate predictions for cell growth and product formation.

With current developments in data analytics and data availability, data driven models can also be applied but these lack process knowledge. Hence, hybrid models combining both mechanistic and data driven model components seem to offer more promise in the move towards digital bioprocessing. An unstructured unsegregated model combined with a data driven (non-parametric) model potentially provides a better performing dynamic model with the ability to extrapolate, whilst also being able to incorporate variations in the culture environment.

New hybrid models can potentially reduce the costs associated with process development in biopharmaceutical manufacturing. The availability of process data is crucial to the development of such models, however, requiring biopharmaceutical companies and potentially research laboratories to ensure their data are recorded in an easily accessible manner.

The developing of a robust hybrid model requires initial experimental data covering a wide design space, as well as process knowledge and process understanding, to design experiments and model structure. Whilst initial development is resource intensive, the developed model can potentially be used in process development, to monitor cell culture performance and to control suitable PAT technologies. The transferability of models, however, from one design space to another, such as between cell lines, products, scales or media composition needs careful consideration and potentially further research and development.

With appropriate process knowledge, process data and the input of new developments in data intensive computational approaches, hybrid models can potentially be developed for use in process development, cell culture monitoring and at a later stage in process control, supporting the move towards Biopharma 4.0.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was financially supported by the Victorian Government under the Higher Education State Investment Fund (VHESIF) program with co-funding from CSL Limited, Australia.

References

Aehle, M., Bork, K., Schaepe, S., Kuprijanov, A., Horstkorte, R., Simutis, R., Lübbert, A, 2012. Increasing batch-to-batch reproducibility of CHO-cell cultures using a model predictive control approach. Cytotechnology 64, 623–634.
Aghili, Z.S., Zarkesh-Esfahani, S.H., 2018. Taguchi experimental design for optimization of recombinant human growth hormone production in CHO cell lines and comparing its biological activity with prokaryotic growth hormone. Drug research 68, 80–88.
Alavijeh, M.K., Baker, I., Lee, Y.Y., Gras, S.L., 2022. Digitally enabled approaches for the scale up of mammalian cell bioreactors. Digital Chemical Engineering, 100040.
Alhuthali, S., Kotidis, P., Kontoravdi, C., 2021. Osmolality effects on CHO cell growth, cell volume, antibody productivity and glycosylation. International Journal of Molecular Sciences 22, 3290.
Amribt, Z., Niu, H., Bogaerts, P., 2013. Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures. Biochemical engineering journal 70, 196–209.
Antonakoudis, A., Strain, B., Barbosa, R., Del Val, I.J., Kontoravdi, C., 2021. Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells. Computers & Chemical Engineering 154, 107471.
Baker, S.M., Poskar, C.H., Schreiber, F., Junker, B.H., 2015. A unified framework for estimating parameters of kinetic biological models. BMC bioinformatics 16, 1–21.
Balcerczyk, A., Damblon, C., Elena-Herrmann, B., Panthu, B., Rautureau, G.J., 2020. Metabolomic approaches to study chemical exposure-related metabolism alterations in mammalian cell cultures. International Journal of Molecular Sciences 21, 6843.
Bandara, S., Schlöder, J.P., Eils, R., Bock, H.G., Meyer, T, 2009. Optimal experimental design for parameter estimation of a cell signaling model. PLoS computational biology 5, e1000558.
Batt, B.C., Kompala, D.S., 1989. A structured kinetic modeling framework for the dynamics of hybridoma growth and monoclonal antibody production in continuous suspension cultures. Biotechnology and bioengineering 34, 515–531.
Bayer, B., Dalmau Diaz, R., Melcher, M., Striedner, G., Duerkop, M., 2021a. Digital twin application for model-based doe to rapidly identify ideal process conditions for space-time yield optimization. Processes 9, 1109.
Bayer, B., Duerkop, M., Striedner, G., Sissolak, B., 2021b. Model transferability and reduced experimental burden in cell culture process development facilitated by hybrid modeling and intensified design of experiments. Frontiers in Bioengineering and Biotechnology 9.
Bayer, B., Sissolak, B., Duerkop, M., Von Stosch, M., Striedner, G., 2020a. The shortcomings of accurate rate estimations in cultivation processes and a solution for precise and robust process modeling. Bioprocess and biosystems engineering 43, 169–178.
Bayer, B., Striedner, G., Duerkop, M., 2020b. Hybrid modeling and intensified DoE: an approach to accelerate upstream process characterization. Biotechnology Journal 15, 2000121.
Bayer, B., Von Stosch, M., Striedner, G., Duerkop, M., 2020c. Comparison of modeling methods for DoE-based holistic upstream process characterization. Biotechnology journal 15, 1900551.
Bayrak, E.S., Wang, T., Cinar, A., Undey, C., 2015. Computational modeling of fed-batch cell culture bioreactor: hybrid agent-based approach. Ifac-Papersonline 48, 1252–1257.
Ben Yahia, B., Malphettes, L., Heinzle, E., 2015. Macroscopic modeling of mammalian cell growth and metabolism. Applied microbiology and biotechnology 99, 7009–7024.
Bree, M.A., Dhurjati, P., Geoghegan Jr, R.F., Robnett, B., 1988. Kinetic modelling of hybridoma cell growth and immunoglobulin production in a large-scale suspension culture. Biotechnology and bioengineering 32, 1067–1072.
Brinc, M., Belič, A, 2019. Optimization of process conditions for mammalian fed-batch cell culture in automated micro-bioreactor system using genetic algorithm. Journal of biotechnology 300, 40–47.
Bui, L., Joswiak, M., Castillo, I., Phillips, A., Yang, J., Hickman, D., 2021. A Hybrid Modeling Approach for Catalyst Monitoring and Lifetime Prediction. ACS Engineering Au 2, 17–26.
Cabaneros Lopez, P., Udugama, I.A., Thomsen, S.T., Roslander, C., Junicke, H., Iglesias, M.M., Gernaey, K.V., 2021. Transforming data to information: A parallel hybrid model for real-time state estimation in lignocellulosic ethanol fermentation. Biotechnology and bioengineering 118, 579–591.
Carvalho, M., Riesberg, J., Budman, H., 2022. Hybrid model to predict the effect of complex media changes in mammalian cell cultures. Biochemical Engineering Journal 186, 108560.
Charaniya, S., Le, H., Rangwala, H., Mills, K., Johnson, K., Karypis, G., Hu, W.-S., 2010. Mining manufacturing data for discovery of high productivity process characteristics. Journal of biotechnology 147, 186–197.
Chen, N., Bennett, M.H., Kontoravdi, C., 2014. Analysis of Chinese hamster ovary cell metabolism through a combined computational and experimental approach. Cytotechnology 66, 945–966.
Coulet, M., Kepp, O., Kroemer, G., Basmaciogullari, S., 2022. Metabolic Profiling of CHO Cells during the Production of Biotherapeutics. Cells 11, 1929.
Craven, S., Shirsat, N., Whelan, J., Glennon, B., 2013. Process model comparison and transferability across bioreactor scales and modes of operation for a mammalian cell bioprocess. Biotechnology Progress 29, 186–196.
De Alwis, D.M., Dutton, R.L., Scharer, J., Moo-Young, M., 2007. Statistical methods in media optimization for batch and fed-batch animal cell culture. Bioprocess and biosystems engineering 30, 107–113.

Del Val, I.J., Kyriakopoulos, S., Albrecht, S., Stockmann, H., Rudd, P.M., Polizzi, K.M., Kontoravdi, C., 2021. CHOmpact: a reduced metabolic model of Chinese hamster ovary cells with enhanced interpretability. bioRxiv.
Dong, Y., Fan, Q., Yan, X., Guo, M., Lu, F., 2014. Development of a hybrid model for sodium gluconate fermentation by Aspergillus niger. Journal of Chemical Technology & Biotechnology 89, 1875–1882.
Endrenyi, L., Tang, H.-Y., 1980. Robust parameter estimation for a simple kinetic model. Computers and Biomedical Research 13, 430–436.
Erdem, C., Birtwistle, M.R., 2023. MEMMAL: A tool for expanding large-scale mechanistic models with machine learned associations and big datasets. Frontiers in Systems Biology 3, 1099413.
Farzan, P., Ierapetritou, M.G, 2018. A framework for the development of integrated and computationally feasible models of large-scale mammalian cell bioreactors. Processes 6, 82.
Farzan, P., Mistry, B., Ierapetritou, M.G., 2017. Review of the important challenges and opportunities related to modeling of mammalian cell bioreactors. AIChE Journal 63, 398–408.
Fernandes, S., Richelle, A., Amribt, Z., Dewasme, L., Bogaerts, P., Wouwer, A.V., 2015. Extended and Unscented Kalman Filter design for hybridoma cell fed-batch and continuous cultures. IFAC-PapersOnLine 48, 1108–1113.
Fernandez-De-Cossio-Diaz, J., Leon, K., Mulet, R, 2017. Characterizing steady states of genome-scale metabolic networks in continuous cell cultures. PLoS computational biology 13, e1005835.
Fisher, O.J., Watson, N.J., Escrig, J.E., Witt, R., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L., 2020. Considerations, challenges and opportunities when developing data-driven models for process manufacturing systems. Computers & Chemical Engineering 140, 106881.
Fisher, O.J., Watson, N.J., Porcu, L., Bacon, D., Rigley, M., Gomes, R.L., 2022. Data-driven modelling for resource recovery: Data volume, variability, and visualisation for an industrial bioprocess. Biochemical Engineering Journal 185, 108499.
Gadgil, M., 2015. Development of a mathematical model for animal cell culture without pH control and its application for evaluation of clone screening outcomes in shake flask culture. Journal of Chemical Technology & Biotechnology 90, 166–175.
Ghorbaniaghdam, A., Chen, J., Henry, O., Jolicoeur, M., 2014. Analyzing clonal variation of monoclonal antibody-producing CHO cell lines using an in silico metabolomic platform. PloS one 9, e90832.
Ghorbaniaghdam, A., Henry, O., Jolicoeur, M., 2013. A kinetic-metabolic model based on cell energetic state: study of CHO cell behavior under Na-butyrate stimulation. Bioprocess and biosystems engineering 36, 469–487.
Glacken, M., Huang, C., Sinskey, A., 1989. Mathematical descriptions of hybridoma culture kinetics. III. Simulation of fed-batch bioreactors. Journal of biotechnology 10, 39–65.
Goudar, C.T., 2012. Computer programs for modeling mammalian cell batch and fed-batch cultures using logistic equations. Cytotechnology 64, 465–475.
Graefe, J., Bogaerts, P., Castillo, J., Cherlet, M., Wérenne, J., Marenbach, P., Hanus, R, 1999. A new training method for hybrid models of bioprocesses. Bioprocess Engineering 21, 423–429.
Grilo, A.L., Mantalaris, A., 2019. A Predictive Mathematical Model of Cell Cycle, Metabolism, and Apoptosis of Monoclonal Antibody-Producing GS–NS0 Cells. Biotechnology Journal 14, 1800573.
Guresen, E., Kayakutlu, G., 2011. Definition of artificial neural networks with comparison to other networks. Procedia Computer Science 3, 426–433.
Gutierrez, J.M., Feizi, A., Li, S., Kallehauge, T.B., Hefzi, H., Grav, L.M., Ley, D., Baycin Hizal, D., Betenbaugh, M.J., Voldborg, B., 2020. Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. Nature communications 11, 1–10.
Harrison, J.U., Baker, R.E., 2018. The impact of temporal sampling resolution on parameter inference for biological transport models. PLoS computational biology 14, e1006235.
Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., 2016. A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. Cell Systems 3, 434–443 e8.
Huang, Z., Xu, J., Yongky, A., Morris, C.S., Polanco, A.L., Reily, M., Borys, M.C., Li, Z.J., Yoon, S., 2020. CHO cell productivity improvement by genome-scale modeling and pathway analysis: Application to feed supplements. Biochemical Engineering Journal 160, 107638.
Hutter, C., Von Stosch, M., Cruz Bournazou, M.N., Butté, A, 2021. Knowledge transfer across cell lines using hybrid Gaussian process models with entity embedding vectors. Biotechnology and Bioengineering 118, 4389–4401.
Jandt, U., Platas Barradas, O., Pörtner, R., Zeng, A.P, 2015. Synchronized mammalian cell culture: Part II—population ensemble modeling and analysis for development of reproducible processes. Biotechnology progress 31, 175–185.
Jang, J.D., Barford, J.P., 2000. An unstructured kinetic model of macromolecular metabolism in batch and fed-batch cultures of hybridoma cells producing monoclonal antibody. Biochemical Engineering Journal 4, 153–168.
Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374, 20150202.
Kahrs, O., Marquardt, W., 2008. Incremental identification of hybrid process models. Computers & Chemical Engineering 32, 694–705.
Kargi, F., 2009. Re-interpretation of the logistic equation for batch microbial growth in relation to Monod kinetics. Letters in applied microbiology 48, 398–401.
Karra, S., Sager, B., Karim, M.N., 2010. Multi-scale modeling of heterogeneities in mammalian cell culture processes. Industrial & Engineering Chemistry Research 49, 7990–8006.
Kasbawati, Samsir, Sulfahri, R., Jaya, A.K., Kalondeng, A, 2018. Determining an appropriate unstructured kinetic model for batch ethanol fermentation data using a direct search method. Biotechnology & Biotechnological Equipment 32, 1167–1173.
Kiparissides, A., Pistikopoulos, E., Mantalaris, A., 2015. On the model-based optimization of secreting mammalian cell (GS-NS0) cultures. Biotechnology and bioengineering 112, 536–548.
Kontoravdi, C., Asprey, S.P., Pistikopoulos, E.N., Mantalaris, A., 2005. Application of global sensitivity analysis to determine goals for design of experiments: an example study on antibody-producing cell cultures. Biotechnology progress 21, 1128–1135.
Kontoravdi, C., Asprey, S.P., Pistikopoulos, E.N., Mantalaris, A., 2007. Development of a dynamic model of monoclonal antibody production and glycosylation for product quality monitoring. Computers & chemical engineering 31, 392–400.
Kontoravdi, C., Pistikopoulos, E.N., Mantalaris, A., 2010. Systematic development of predictive mathematical models for animal cell cultures. Computers & Chemical Engineering 34, 1192–1198.
Kotidis, P., Jedrzejewski, P., Sou, S.N., Sellick, C., Polizzi, K., Del Val, I.J., Kontoravdi, C., 2019. Model-based optimization of antibody galactosylation in CHO cell culture. Biotechnology and bioengineering 116, 1612–1626.
Kotidis, P., Kontoravdi, C., 2020. Harnessing the potential of artificial neural networks for predicting protein glycosylation. Metabolic engineering communications 10, e00131.
Kotidis, P., Pappas, I., Avraamidou, S., Pistikopoulos, E.N., Kontoravdi, C., Papathanasiou, M.M., 2021. DigiGlyc: A hybrid tool for reactive scheduling in cell culture systems. Computers & Chemical Engineering 154, 107460.
Kusena, J., Shariatzadeh, M., Thomas, R., Wilson, S., 2021a. Understanding cell culture dynamics: a tool for defining protocol parameters for improved processes and efficient manufacturing using human embryonic stem cells. Bioengineered 12, 979–996.
Kusena, J.W.T., Shariatzadeh, M., Studd, A.J., James, J.R., Thomas, R.J., Wilson, S.L., 2021b. The importance of cell culture parameter standardization: an assessment of the robustness of the 2102Ep reference cell line. Bioengineered 12, 341–357.
Kyriakopoulos, S., Ang, K.S., Lakshmanan, M., Huang, Z., Yoon, S., Gunawan, R., Lee, D. Y., 2018. Kinetic modeling of mammalian cell culture bioprocessing: the quest to advance biomanufacturing. Biotechnology journal 13, 1700229.
Laursen, S.Ö., Webb, D., Ramirez, W.F., 2007. Dynamic hybrid neural network model of an industrial fed-batch fermentation process to produce foreign protein. Computers & chemical engineering 31, 163–170.
Le Guennec, A., Tea, I., Antheaume, I., Martineau, E., Charrier, B., Pathan, M., Akoka, S., Giraudeau, P., 2012. Fast determination of absolute metabolite concentrations by spatially encoded 2D NMR: application to breast cancer cell extracts. Analytical chemistry 84, 10831–10837.
Le, H., Kabbur, S., Pollastrini, L., Sun, Z., Mills, K., Johnson, K., Karypis, G., Hu, W.-S., 2012. Multivariate analysis of cell culture bioprocess data—lactate consumption as process indicator. Journal of biotechnology 162, 210–223.
Lee, D., Jayaraman, A., Kwon, J.S., 2020. Development of a hybrid model for a partially known intracellular signaling pathway through correction term estimation and neural network modeling. PLoS Computational Biology 16, e1008472.
Li, F., Vijayasankaran, N., Shen, A., Kiss, R., Amanullah, A., 2010. Cell culture processes for monoclonal antibody production. MAbs. Taylor & Francis 466–479.
Linardos, T.I., Kalogerakis, N., Behie, L.A., 1992. Cell cycle model for growth rate and death rate in continuous suspension hybridoma cultures. Biotechnology and bioengineering 40, 359–368.
Liu, Y.-H., Bi, J.-X., Zeng, A.-P., Yuan, J.-Q., 2008. A simple kinetic model for myeloma cell culture with consideration of lysine limitation. Bioprocess and biosystems engineering 31, 569–577.
Löpez-Meza, J., Araíz-Hernández, D., Carrillo-Cocom, L.M., López-Pacheco, F., Rocha-Pizaña, M.D.R., Alvarez, M.M, 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.
Luna, M.F., Ochsner, A.M., Amstutz, V., Von Blarer, D., Sokolov, M., Arosio, P., Zinn, M., 2021. Modeling of continuous PHA production by a hybrid approach based on first principles and machine learning. Processes 9, 1560.
Luo, Y., Kurian, V., Ogunnaike, B.A., 2021. Bioprocess systems analysis, modeling, estimation, and control. Current Opinion in Chemical Engineering 33, 100705.
Mantzaris, N.V., 2006. Stochastic and deterministic simulations of heterogeneous cell population dynamics. Journal of theoretical biology 241, 690–706.
Martínez, J.A., Bulté, D.B., Contreras, M.A., Palomares, L.A., Ramírez, O.T, 2020. Dynamic modeling of CHO cell metabolism using the hybrid cybernetic approach with a novel elementary mode analysis strategy. Frontiers in bioengineering and biotechnology 8, 279.
Mears, L., Stocks, S.M., Albaek, M.O., Sin, G., Gernaey, K.V., 2017. Mechanistic fermentation models for process design, monitoring, and control. Trends in biotechnology 35, 914–924.
Merkelbach, K., Schweidtmann, A.M., Müller, Y., Schwoebel, P., Mhamdi, A., Mitsos, A., Schuppert, A., Mrziglod, T., Schneckener, S, 2022. HybridML: Open source platform for hybrid modeling. Computers & Chemical Engineering 160, 107736.
Meshram, M., Naderi, S., Mcconkey, B., Budman, H., Scharer, J., Ingalls, B., 2012. Population-based modeling of the progression of apoptosis in mammalian cell culture. Biotechnology and Bioengineering 109, 1193–1204.
Möller, J., Korte, K., Pörtner, R., Zeng, A.-P., Jandt, U, 2020. Near-Physiological Cell Cycle Synchronization with Countercurrent Centrifugal Elutriation. Animal Cell Biotechnology.
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. Biotechnology and bioengineering 115, 2996–3008.

Möller, J., Kuchemüller, K.B., Steinmetz, T., Koopmann, K.S., Pörtner, R, 2019. Model-assisted design of experiments as a concept for knowledge-based bioprocess development. Bioprocess and biosystems engineering 42, 867–882.
Moreno, M.S., López, E., Divins, N.J., Llorca, J, 2014. Parameter estimation of an empirical kinetic model for CO preferential oxidation. XLIII Jornadas Argentinas de Informática e Investigación Operativa (43JAIIO)-XII Simposio Argentino de Investigación Operativa (SIO)(Buenos Aires 2014.
Naderi, S., Meshram, M., Wei, C., Mcconkey, B., Ingalls, B., Budman, H., Scharer, J., 2011. Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells. Biotechnology progress 27, 1197–1205.
Narayanan, H., Behle, L., Luna, M.F., Sokolov, M., Guillén-Gosálbez, G., Morbidelli, M., Butté, A, 2020. Hybrid-EKF: Hybrid Model coupled with Extended Kalman Filter for real-time monitoring and control of mammalian cell culture. Biotechnology and Bioengineering 117, 2703–2714.
Narayanan, H., Luna, M., Sokolov, M., Butte, A., Morbidelli, M., 2022. Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Cell Culture Processes. Industrial & Engineering Chemistry Research 61, 8658–8672.
Narayanan, H., Sokolov, M., Morbidelli, M., Butté, A, 2019. A new generation of predictive models: the added value of hybrid models for manufacturing processes of therapeutic proteins. Biotechnology and Bioengineering 116, 2540–2549.
Nargund, S., Guenther, K., Mauch, K., 2019. The move toward biopharma 4.0: insilico biotechnology develops “smart” processes that benefit biomanufacturing through digital twins. Genetic Engineering & Biotechnology News 39, 53–55.
Nickel, D.B., Cruz-Bournazou, M.N., Wilms, T., Neubauer, P., Knepper, A., 2017. Online bioprocess data generation, analysis, and optimization for parallel fed-batch fermentations in milliliter scale. Wiley Online Library.
Nikita, S., Thakur, G., Jesubalan, N.G., Kulkarni, A., Yezhuvath, V.B., Rathore, A.S., 2022. AI-ML applications in bioprocessing: ML as an enabler of real time quality prediction in continuous manufacturing of mAbs. Computers & Chemical Engineering 164, 107896.
Nolan, R.P., Lee, K., 2011. Dynamic model of CHO cell metabolism. Metabolic engineering 13, 108–124.
Nold, V., Junghans, L., Bisgen, L., Drerup, R., Presser, B., Gorr, I., Schwab, T., Knapp, B., Wieschalka, S., 2021. Applying intensified design of experiments to mammalian cell culture processes. Engineering in Life Sciences.
Noll, P., Henkel, M., 2020. History and evolution of modeling in biotechnology: modeling & simulation, application and hardware performance. Computational and Structural Biotechnology Journal 18, 3309–3323.
O’Brien, C.M., Zhang, Q., Daoutidis, P., Hu, W.-S., 2021. A hybrid mechanistic-empirical model for in silico mammalian cell bioprocess simulation. Metabolic Engineering 66, 31–40.
O’Flaherty, R., Bergin, A., Flampouri, E., Mota, L.M., Obaidi, I., Quigley, A., Xie, Y., Butler, M., 2020. Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. Biotechnology advances 43, 107552.
Oliveira, R., 2004. Combining first principles modelling and artificial neural networks: a general framework. Computers & Chemical Engineering 28, 755–766.
Papathanasiou, M.M., Quiroga-Campano, A.L., Steinebach, F., Elviro, M., Mantalaris, A., Pistikopoulos, E.N., 2017. Advanced model-based control strategies for the intensification of upstream and downstream processing in mAb production. Biotechnology progress 33, 966–988.
Pappenreiter, M., Bayer, B., Logarušić, M., Sissolak, B., Jungbauer, A, 2023. Irreversible and reversible impact on cellular behavior upon intra-experimental process parameter shifts in a CHO semi-continuous perfusion process. Biochemical Engineering Journal 193, 108876.
Park, J.M., Kim, T.Y., Lee, S.Y., 2010. Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses. Proceedings of the National Academy of Sciences 107, 14931–14936.
Park, S.-Y., Park, C.-H., Choi, D.-H., Hong, J.K., Lee, D.-Y., 2021. Bioprocess digital twins of mammalian cell culture for advanced biomanufacturing. Current Opinion in Chemical Engineering 33, 100702.
Pereira, S., Kildegaard, H.F., Andersen, M.R., 2018. Impact of CHO metabolism on cell growth and protein production: an overview of toxic and inhibiting metabolites and nutrients. Biotechnology Journal 13, 1700499.
Pérez-Fernández, B.A., Fernandez-De-Cossio-Diaz, J., Boggiano, T., Leon, K., Mulet, R, 2021. In-silico media optimization for continuous cultures using genome scale metabolic networks: The case of CHO-K1. Biotechnology and Bioengineering 118, 1884–1897.
Pham, T.D., Manapragada, C., Rajan, N., Aickelin, U., 2023. Pharmaceutical process optimisation: Decision support under high uncertainty. Computers & Chemical Engineering 170, 108100.
Pinto, J., Mestre, M., Ramos, J., Costa, R.S., Striedner, G., Oliveira, R., 2022. A general deep hybrid model for bioreactor systems: Combining first principles with deep neural networks. Computers & Chemical Engineering 165, 107952.
Powers, D.N., Trunfio, N., Velugula-Yellela, S.R., Angart, P., Faustino, A., Agarabi, C., 2020. Multivariate data analysis of growth medium trends affecting antibody glycosylation. Biotechnology Progress 36, e2903.
Psichogios, D.C., Ungar, L.H., 1992. A hybrid neural network-first principles approach to process modeling. AIChE Journal 38, 1499–1511.
Puente-Massaguer, E., Badiella, L., Gutiérrez-Granados, S., Cervera, L., Gòdia, F, 2019. A statistical approach to improve compound screening in cell culture media. Engineering in life sciences 19, 315–327.
Radhakrishnan, D., Wells, E.A., Robinson, A.S., 2018. Strategies to enhance productivity and modify product quality in therapeutic proteins. Current opinion in chemical engineering 22, 81–88.
Ram, Y., Dellus-Gur, E., Bibi, M., Karkare, K., Obolski, U., Feldman, M.W., Cooper, T.F., Berman, J., Hadany, L., 2019. Predicting microbial growth in a mixed culture from growth curve data. Proceedings of the National Academy of Sciences 116, 14698–14707.
Ramirez, O.T., Mutharasan, R., 1990. Cell cycle-and growth phase-dependent variations in size distribution, antibody productivity, and oxygen demand in hybridoma cultures. Biotechnology and Bioengineering 36, 839–848.
Ramos, J.R., Oliveira, G.P., Dumas, P., Oliveira, R., 2022. Genome-scale modeling of Chinese hamster ovary cells by hybrid semi-parametric flux balance analysis. Bioprocess and Biosystems Engineering 45, 1889–1904.
Ramsay, J.O., Hooker, G., Campbell, D., Cao, J., 2007. Parameter estimation for differential equations: a generalized smoothing approach. Journal of the Royal Statistical Society Series B: Statistical Methodology 69, 741–796.
Rath, A.G., Rehberg, M., Janke, R., Genzel, Y., Scholz, S., Noll, T., Rose, T., Sandig, V., Reichl, U., 2014. The influence of cell growth and enzyme activity changes on intracellular metabolite dynamics in AGE1. HN. AAT cells. Journal of Biotechnology 178, 43–53.
Rathore, A., Singh, S., 2015. Use of multivariate data analysis in bioprocessing. BioPharm Int 28, 26–31.
Retamal, C., Dewasme, L., Hantson, A.-L., Wouwer, A.V., 2018. Parameter estimation of a dynamic model of Escherichia coli fed-batch cultures. Biochemical Engineering Journal 135, 22–35.
Reyes, S.J., Durocher, Y., Pham, P.L., Henry, O., 2022. Modern sensor tools and techniques for monitoring, controlling, and improving cell culture processes. Processes 10, 189.
Richelle, A., Corbett, B., Agarwal, P., Vernersson, A., Trygg, J., Mccready, C., 2022. Model-based intensification of CHO cell cultures: one-step strategy from fed-batch to perfusion. bioRxiv.
Robitaille, J., Chen, J., Jolicoeur, M., 2015. A single dynamic metabolic model can describe mAb producing CHO cell batch and fed-batch cultures on different culture media. PloS one 10, e0136815.
Roubos, J., Krabben, P., Setnes, M., Babuska, R., Heijnen, J., Verbruggen, H., 1999. Hybrid model development for fed-batch bioprocesses; combining physical equations with the metabolic network and black-box kinetics. 6th Workshop on fuzzy systems, September. Citeseer 8–9.
Salim, T., Chauhan, G., Templeton, N., Ling, W.L.W., 2022. Using MVDA with stoichiometric balances to optimize amino acid concentrations in chemically defined CHO cell culture medium for improved culture performance. Biotechnology and Bioengineering 119, 452–469.
Saraiva, I., Vande Wouwer, A., Hantson, A.-L., 2015. Parameter identification of a dynamic model of CHO cell cultures: an experimental case study. Bioprocess and biosystems engineering 38, 2231–2248.
Schinn, S.-M., Morrison, C., Wei, W., Zhang, L., Lewis, N.E., 2021a. Systematic evaluation of parameters for genome-scale metabolic models of cultured mammalian cells. Metabolic Engineering 66, 21–30.
Schinn, S.M., Morrison, C., Wei, W., Zhang, L., Lewis, N.E., 2021b. A genome-scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures. Biotechnology and Bioengineering 118, 2118–2123.
Schubert, J., Simutis, R., Dors, M., Havlik, I., Lübbert, A, 1994. Bioprocess optimization and control: Application of hybrid modelling. Journal of biotechnology 35, 51–68.
Selişteanu, D., Şendrescu, D., Georgeanu, V., Roman, M, 2015. Mammalian cell culture process for monoclonal antibody production: nonlinear modelling and parameter estimation. BioMed research international 2015.
Selvarasu, S., Ho, Y.S., Chong, W.P., Wong, N.S., Yusufi, F.N., Lee, Y.Y., Yap, M.G., Lee, D.Y., 2012. Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. Biotechnology and bioengineering 109, 1415–1429.
Sha, S., Huang, Z., Wang, Z., Yoon, S., 2018. Mechanistic modeling and applications for CHO cell culture development and production. Current opinion in chemical engineering 22, 54–61.
Shirsat, N., Mohd, A., Whelan, J., English, N.J., Glennon, B., Al-Rubeai, M., 2015a. Revisiting Verhulst and Monod models: analysis of batch and fed-batch cultures. Cytotechnology 67, 515–530.
Shirsat, N.P., English, N.J., Glennon, B., Al-Rubeai, M., 2015b. Modelling of mammalian cell cultures. Animal cell culture 259–326.
Sidoli, F.R., Mantalaris, A., Asprey, S., 2004. Modelling of mammalian cells and cell culture processes. Cytotechnology 44, 27–46.
Slininger, P., Dien, B., Lomont, J., Bothast, R., Ladisch, M., Okos, M., 2014. Evaluation of a kinetic model for computer simulation of growth and fermentation by Scheffersomyces (Pichia) stipitis fed D-xylose. Biotechnology and bioengineering 111, 1532–1540.
Smiatek, J., Jung, A., Bluhmki, E., 2020. Towards a digital bioprocess replica: computational approaches in biopharmaceutical development and manufacturing. Trends in Biotechnology 38, 1141–1153.
Solle, D., Hitzmann, B., Herwig, C., Pereira Remelhe, M., Ulonska, S., Wuerth, L., Prata, A., Steckenreiter, T., 2017. Between the poles of data-driven and mechanistic modeling for process operation. Chemie Ingenieur Technik 89, 542–561.
Sommeregger, W., Sissolak, B., Kandra, K., Von Stosch, M., Mayer, M., Striedner, G., 2017. Quality by control: Towards model predictive control of mammalian cell culture bioprocesses. Biotechnology Journal 12, 1600546.
Sun, Y., Nathan-Roberts, W., Pham, T. D., Otte, E. & Aickelin, U. 2022. Multi-fidelity Gaussian Process for Biomanufacturing Process Modeling with Small Data. arXiv preprint arXiv:2211.14493.
Suzuki, E., Ollis, D., 1989. Cell cycle model for antibody production kinetics. Biotechnology and bioengineering 34, 1398–1402.

Thompson, M.L., Kramer, M.A., 1994. Modeling chemical processes using prior knowledge and neural networks. AIChE Journal 40, 1328–1340.
Venkatasubramanian, V., Mann, V., 2022. Artificial intelligence in reaction prediction and chemical synthesis. Current Opinion in Chemical Engineering 36, 100749.
Villiger, T.K., Scibona, E., Stettler, M., Broly, H., Morbidelli, M., Soos, M., 2016. Controlling the time evolution of mAb N-linked glycosylation-Part II: Model-based predictions. Biotechnology progress 32, 1135–1148.
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 and biosystems engineering 39, 773–784.
Von Stosch, M., Oliveira, R., Peres, J., De Azevedo, S.F, 2014. Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Computers & Chemical Engineering 60, 86–101.
Wang, S., Fan, K., Luo, N., Cao, Y., Wu, F., Zhang, C., Heller, K.A., You, L., 2019. Massive computational acceleration by using neural networks to emulate mechanism-based biological models. Nature communications 10, 4354.
Wang, Z., Wang, C., Chen, G., 2022. Kinetic modeling: A tool for temperature shift and feeding optimization in cell culture process development. Protein Expression and Purification, 106130.
Wechselberger, P., Sagmeister, P., Herwig, C., 2013. Model-based analysis on the extractability of information from data in dynamic fed-batch experiments. Biotechnology progress 29, 285–296.
Xing, Z., Bishop, N., Leister, K., Li, Z.J., 2010. Modeling kinetics of a large-scale fed-batch CHO cell culture by Markov chain Monte Carlo method. Biotechnology progress 26, 208–219.
Xu, J., Tang, P., Yongky, A., Drew, B., Borys, M.C., Liu, S., Li, Z.J., 2019. Systematic development of temperature shift strategies for Chinese hamster ovary cells based on short duration cultures and kinetic modeling. MAbs 191–204.
Yahia, B.B., Malphettes, L., Heinzle, E., 2021. Predictive macroscopic modeling of cell growth, metabolism and monoclonal antibody production: Case study of a CHO fed-batch production. Metabolic Engineering 66, 204–216.
Yeo, H.C., Hong, J., Lakshmanan, M., Lee, D.-Y., 2020. Enzyme capacity-based genome scale modelling of CHO cells. Metabolic engineering 60, 138–147.
Zeng, A.P., 1995. A kinetic model for product formation of microbial and mammalian cells. Biotechnology and bioengineering 46, 314–324.
Zhang, A., Sun, H., Xu, H., Qiu, S., Wang, X., 2013a. Cell metabolomics. Omics: a journal of integrative biology 17, 495–501.
Zhang, H., Wang, H., Liu, M., Zhang, T., Zhang, J., Wang, X., Xiang, W., 2013b. Rational development of a serum-free medium and fed-batch process for a GS-CHO cell line expressing recombinant antibody. Cytotechnology 65, 363–378.
Zhang, J., Petersen, S.D., Radivojevic, T., Ramirez, A., PÉrez-ManrÍquez, A., Abeliuk, E., Sánchez, B.J., Costello, Z., Chen, Y., Fero, M.J, 2020. Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism. Nature communications 11, 1–13.