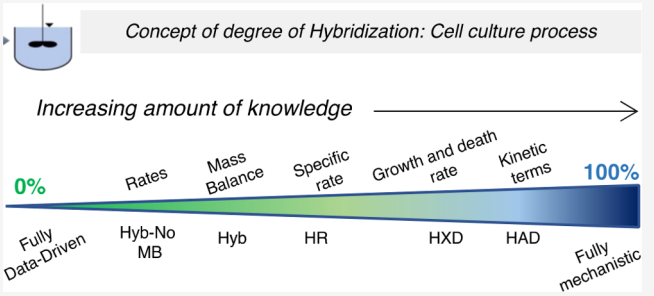
I&EC research
Industrial & Engineering Chemistry Research
pubs.acs.org/IECR
Article

Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Cell Culture Processes

Harini Narayanan, Martin Luna, Michael Sokolov, Alessandro Butté,* and Massimo Morbidelli*

Cite This: *Ind. Eng. Chem. Res.* 2022, 61, 8658−8672
Read Online

ABSTRACT: In this work, we aim to introduce the concept of the degree of hybridization for cell culture process modeling. We propose that a family of hybrid models can be created with varying fractions of process knowledge explicitly encoded in the model, defined as the degree of hybridization, with the two extremes being fully data-driven (0%) and fully mechanistic (100%) models. Subsequently, the aim is to compare the different models based on different metrics: model accuracy, the experimental effort for model development, extrapolation capability, the capability of generating new process understanding, and ease of utilization in practice, and to demonstrate that this could provide an additional degree of freedom for model selection. We could quantitatively demonstrate that for the cell culture process, either extreme has limitations. The major drawback of the data-driven model is the poor performance at low data availability as well as poor extrapolation capability, inability to provide process understanding, and subsequently inefficient practical application. On the other hand, the mechanistic model has poor accuracy due to the addition of excessive knowledge that then biases the models. Moving from data-driven to mechanistic models, the performance of the models improves progressively, as long as the knowledge added is not too biased. We show that the choice of the hybrid models to be used is based on the goal of model development. For instance, hybrid models including mass balances on each species show better performance in transferring models across different modes of operation. On the other hand, models with a higher degree of hybridization allow for more process interpretation possibilities. For modeling accuracy, amount of training data, extrapolation, and practical applications, the Hybrid Rate (HR) model is found to have the optimal degree of hybridization. This is likely due to the compromise between adding process knowledge and increasing the model parameters achieved by the HR model. The HR model features the incorporation of mass balance and channels the data-driven modeling to cell-specific rates, and thus these two pieces of information appear to be the most crucial ones. Finally, we believe that the concept will be instrumental in progressively developing and testing hypotheses about complex processes such as cell cultures.



1. INTRODUCTION

Mathematical models have traditionally been regarded as a tool, complementary and alternative to experimental investigation, to understand and describe reality. Thus, like in many other fields of science and engineering, modeling has been applied also to biotechnological manufacturing processes. The use of models has become crucial for various process applications such as optimization, monitoring, control, and scale-up.[1] In the last decade in the biopharmaceutical industry, technological possibilities to enhance automation, digitalization,[2,3] and supporting initiatives by the regulatory agencies have further emphasized the importance of mathematical models.[4−6]

Mathematical models are typically categorized into data-driven (DD) (or statistical) and knowledge-driven (or mechanistic) models. The former models are preferred when no or little understanding about the model input−output relationship is available and are alternatively referred to as black-box models. They rely on using statistical methods or machine learning algorithms to infer direct interrelationships between the model input and output variables from suitable sets of experimental data.[7] In contrast, knowledge-based models are preferred when a precise understanding of the detailed mechanisms underlying the relation between the process inputs and outputs is available.[7] Accordingly, the individual physicochemical elementary processes involved in the overall process are represented using suitable mathematical expressions. However, performing ab initio modeling with detailed molecular-level understanding is resource-intensive and thus rarely performed for industrial processes. Subsequently, mechanistic models rely on approximate expressions

Received: December 15, 2021
Revised: April 20, 2022
Accepted: April 21, 2022
Published: June 16, 2022

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

(A)

	IN SILICO DATASET	EXPERIMENTAL
Scale	Unit Volume	Bio Reactor (3.5 L)
Product	mAb	Fusion Fc protein
Medium	Chemically defined	Peptone - based
Mode	Fed - batch	Fed - batch
Culture duration (days)/ T_batch	14	10
Sampling	Once per day	Once per day
Process Condition (Z)	Stirring rate	Amplification duration, Amplification VCD, Amplification cell age
		DO, pH_{set}
Process variables (X)	VCD, GLC, LAC,	VCD, GLC, LAC
	GLN, NH4, Titer	GLN, GLU, NH4, OSM, Titer
Feed Species (F)	GLC, GLN	GLC
Feed Start, End	Designed	Day 3 to Day 9
Controllable variable (W)	pH, T	pH, pCO_2, pO_2
Shifts	pH, T	None
Variables Predicted	Process variables (X)	Process variables (X)
Total # of runs	Sample from emulator	81

(B)

Dataset	Test Sets	# Exp	Sampling
In-Silico	Interpolation	20	Within the design space for all factors
	Extrapolation 1	20	Outside the design space for initial conditions
	Extrapolation 2	20	Outside the design space for feeding strategy – amount and days
	Extrapolation 3	20	Outside the design space for initial VCD
Experimental	Interpolation	16	Stratified, 20% test from each cluster identified in the PCA space
	Extrapolation	51	Different process behavior of cell line

(C)

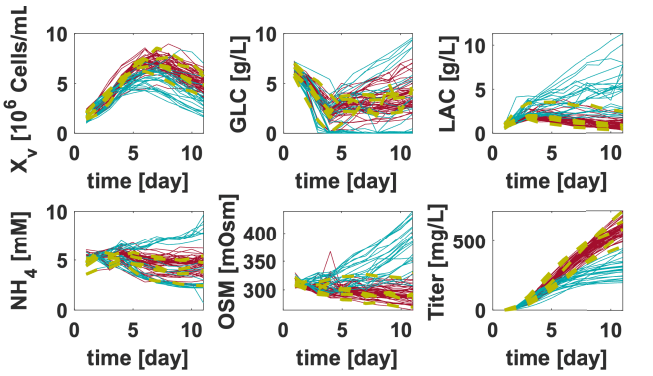


Figure 1. (A) Tabulation of the key characteristics of the in silico as well as the experimental dataset used. (B) Summary of the rationale of train-test split performed for both the in silico and experimental dataset. (C) Dynamic profiles of the key process variables: X_v, glucose (GLC), lactose (LAC), NH_4, osmolality (OSM), and Titer, constituting the experimental data set, selected for training the different models (blue and yellow) and testing (red) their extrapolation capabilities (cf. "extrapolation" set, B).

to describe the underlying phenomena at a molecular level, which may or may not be understood.

Typically, the choice of the modeling approach depends on the availability of process knowledge, historical data, and the capacity to generate ad hoc data. For instance, in the case of chromatographic processes,[8] a considerable amount of mechanistic knowledge (e.g., convection, axial dispersion) and its corresponding mathematical representation is available. The lack of knowledge is limited to a few phenomena, such as effective pore diffusion and the adsorption isotherm. The contrary is experienced in cell culture processes where the mechanistic knowledge of the process is very limited[1,9] or very complex to be applied for industrial applications (for instance, models based on metabolic networks using methods such as flux analysis).[10−16] Thus, macro-kinetic models assuming Monod kinetics for metabolites are typically used as mechanistic models.[17,18]

In this work, the focus is exclusively on the biopharmaceutical cell culture processes, where the state-of-the-art process characterization method is such that simple data-driven models are generally used for process modeling. This is due to two primary reasons: (i) only partial understanding of the elementary physicochemical processes involved in a cell culture process; (ii) practical limitations in the data analytics, making it impossible to develop detailed mechanistic models (e.g., ones based on the metabolic network of the cells); and (iii) the lack of experts capable to develop hybrid approaches. For this reason, the limited complexity of linear data-driven approaches, e.g., based on partial least square (PLS) regression, makes them simple to generate, to use, and eventually to validate in a usually narrow application range. The use of more complex nonlinear data-driven approaches to building general models covering the space of operating conditions is limited by the fact that experiments with cell cultures are very resource-intensive and, therefore, generated data relatively scarce. In this situation, hybrid models, which synergistically combine the advantages of data- and knowledge-driven models, arise naturally as a pragmatic solution to integrate the two approaches.[2,7,9] The concept involves describing the elementary physicochemical phenomena involved in the process, which are well understood through appropriate mechanistic mathematical expressions while leaving the remaining part of the relationships between process input and output variables (often related to the intrinsic mechanisms of the cell) to data-driven approaches.[7,19−23]

Successful application cases include various chemical processes,[24−37] energy storage and management,[38] and several areas of biotechnology[20,21,23,38−43] including fermentations[44] and cell cultures.[9,45,46] Their use has been emphasized and demonstrated for different process applications such as optimization,[47−51] quality modeling,[4] monitoring,[52−54] and control.[50,51,53,54] Subsequently, it is considered as a key future technology for the realization of Quality by Design—Process Analytical Technology (QbD-PAT)[55] and industry 4.0[2−4] initiatives.

However, in all of these previous studies, hybrid models are developed based on an a priori definition of the parts of the model described through a data-driven or a mechanistic approach, and these are assumed constant during the entire model application and analysis. In this work, instead, we generate a family of hybrid models which incorporate process knowledge and engineering know-how to different extents. Here, we reapply the concept of "degree of hybridization" for cell culture processes, introduced and described in our

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

Table 1. Summary of the Different Models along the Hybridization Axis Ordered from 0% (First Row) to 100% (Last Row) with the Fragment of Knowledge Explicitly Added in the Model (in Addition to the Previous Models) and the Corresponding Modeling Equation.

Model	Sequential explicit mechanism added	Equation	Input to DD model
BWU-PLS1			Z, $X(t-1)$, $W(t-1)$, $F(t-1)$
Hyb-No MB	rate of accumulation	$\frac{dX_i}{dt} = NN_i$	Z, $X(t)$, $W(t)$, $F(t)$
Hyb	mass balance	$\frac{dX_i}{dt} = NN_i + In_i - Out_i$	Z, $X(t)$, $W(t)$
HR	specific rate	$\frac{dX_i}{dt} = NN_i \cdot X_v + In_i - Out_i$	Z, X[except X_v](t), $W(t)$
HXD	specific X_v growth and death rate	$\frac{dX_v}{dt} = NN_{1a} \cdot X_v - NN_{1b} \cdot X_v + In_i - Out_i$ $\frac{dX_i}{dt} = NN_i \cdot X_v + In_i - Out_i$	Z, X[except X_v](t), $W(t)$
HAD	kinetic term	$\frac{dX_v}{dt} = NN_{1a} \cdot X_v - NN_{1b} \cdot X_v + In_i - Out_i$ $\frac{dX_i}{dt} = NN_{ia} \cdot X_v + NN_{ib} + In_i - Out_i$	Z, X[except X_v](t), $W(t)$
Mech	Monod equation for metabolites	$\frac{dX_v}{dt} = (\mu_g - \mu_d)X_v$ $\frac{dGLC}{dt} = -\left(\frac{\mu_g}{Y_{x/glc}} + m_{glc}\right)X_v + In_{Glc} - Out_{Glc}$ $\frac{dGLN}{dt} = -\left(\frac{\mu_g}{Y_{x/gln}} + m_{gln}\right)X_v - d_{gln}GLN + In_{Gln} - Out_{Gln}$ $\frac{dLAC}{dt} = Y_{lac/glc}\left(\frac{\mu_g}{Y_{x/glc}} + m_{glc}\right)X_v$ $\frac{dNH_4}{dt} = Y_{am/gln}\left(\frac{\mu_g}{Y_{x/gln}} + m_{gln}\right)X_v + d_{gln}GL\,N$ $\frac{dTiter}{dt} = q_p X_v$ $\mu_g = \mu_g^{max}\frac{GLC}{GLC + K_{glc}}\frac{GLN}{GLN + K_{gln}}\frac{KI_{lac}}{LAC + KI_{lac}}\frac{KI_{am}}{NH_4 + KI_{am}}$ $\mu_d = \mu_d^{max}\frac{LAC}{LAC + KD_{lac}}\frac{NH_4}{NH_4 + KD_{am}}$ $m_{gln} = a_{gln1}\frac{GLN}{GLN + a_{gln2}}$	

previous work on biochromatography.[56] Essentially, the degree of hybridization is a qualitative indicator to describe the extent of process knowledge incorporated in a given hybrid model.[56] In particular, we define 0% hybridized models as the traditional data-driven models, and, on the other hand, 100% hybridized models as the traditional mechanistic models. The term hybrid models or family of hybrid models denotes, therefore, models with intermediate degrees of hybridization. The scope of this analysis is to study how the incorporation of different amounts of mechanistic knowledge in the model can impact metrics like model accuracy, the experimental effort for parameter estimation, extrapolation capability, the capability of generating new process understanding, and ease of utilization in practice.

2. MATERIALS AND METHODS

2.1. Data Sets. The design factors and the data collected during the runs are grouped into different sets in line with process data organization and nomenclature described in our previous works.[9,54,57] In brief, the different information matrices are as follows: (i) Z: process conditions that are kept constant throughout the run, (ii) X: process variables that are dynamically measured (iii) W: controlled variables, and (iv) F: the amount of feed added. Two fed-batch mammalian cell culture datasets (with daily bolus feed), one generated by an in silico macro-kinetic model and the other from an industrial lab-scale bioreactor, are used in this paper to investigate the performance of the considered models. Figure 1A summarizes the variable constituting the different information matrices for both datasets.

2.1.1. In Silico Data Set. The macro-kinetics models for mammalian cell cultures proposed by refs 17 and 18 and appropriately modified in this work to account for the effect on T, pH, and stirring rates on different specific rates, were used to produce the in silico data sets. There are 14 process factors in the in silico model, namely, (i) stirring rate, (ii) the initial conditions: X_v, GLC, and glutamine (GLN), (iii) setpoint of controlled variables: pH before and after shift, temperature before and after shift, days of shift, and (iv) feeding strategy: GLC and GLN feeding start day−end day, and the amount of

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

bolus feed for GLC and GLN, respectively (kept constant for all of the days of feeding). The in silico model provides daily concentration values of six process variables, as summarized in Figure 1A. These are further perturbed with white noise, having a relative standard deviation of 10% with respect to the true simulated value. A detailed description of the dataset and its corresponding organization into the different information matrices can be found in our previous work,[54] with the key characteristics summarized in Figure 1A.

Datasets with different numbers of runs were simulated using the Latin Hyper Cube (LHC) sampling design (uniform space-filling design) and used for model training to demonstrate the minimum data required for respective model development. Subsequently, to analyze the performance of the model, four independent test datasets with different characteristics, as summarized in Figure 1B, were considered. Correspondingly, each test set was sampled using an independent LHC design in the respective design space. The "interpolation" test set is used to analyze the predictive accuracy of the model (cf. Section 3.1), while the "extrapolation 1" and "extrapolation 2" test sets are used to represent the general extrapolation capability of the different models (cf. Section 3.2). The extrapolation 2 and "extrapolation 3" test sets are then used to compare the extrapolation capability among the hybrid models (Section 3.4). Subsequently, we highlight the advantage of introducing relevant fragments of knowledge, namely, the mass balance and the functionality with respect to X_v, in the modeling process.

2.1.2. Experimental Data Set. In addition, the performance of the considered models is evaluated on an experimental dataset obtained from an industrial mammalian cell culture fed-batch bioreactor (3 L scale). Figure 1A summarizes the main characteristic of the dataset. The technical details of the process can be found in the original publication[58] with different data analyses performed in the works.[9,54,59,60] As summarized in Figure 1B, the interpolation test set is obtained through a stratified sampling of 20% of the experiments from each cluster identified in the principal component analysis (PCA) space to ensure coverage of the design space. It is used to compare the predictive accuracy of the different models along the hybridization axis for a dataset of industrial relevance. For the extrapolation performance comparison, the dataset (with 80 experiments) is split into two parts, representing two different process behaviors of the system, identified in our previous data analysis work on this dataset.[61] The process behaviors were observed to be different when DO < 25% (blue profiles in Figure 1C) and DO ≥ 25% (red and yellow profiles in Figure 1C). The former profiles are characterized by very low or very high GLC and NH_4, high OSM, and increasing LAC, which result in low X_v and titer. The latter profiles are characterized by intermediate GLC and NH_4 values, decreasing OSM and consumption of LAC, generally corresponding to higher X_v and titers. In this case study, we train models on the first group, i.e., the blue profiles (24 experiments), with the addition of only five experiments of the second group, i.e., the yellow profiles, and then we test the model performance on the remaining 51 experiments of the second group, i.e., the red profiles. The five experiments (indicated in yellow) are selected such that the pairwise distance between the experiments (computed using the batch-wise unfolded data created using all of the information matrices) is maximized. Here, the motivation is to show how the knowledge created in a system with a certain process behavior (due to a certain metabolic state or type of product) can be used to minimize the data required for a new system.

2.2. Models. We define the degree of hybridization as a qualitative measure that signifies the amount of process knowledge and engineering know-how that is described explicitly in the mechanistic backbone of the hybrid models. In the following, we consider seven possible hybrid models, characterized by a different degree of hybridization ranging from 0% (data-driven) to 100% hybridized models (mechanistic), as summarized in Table 1. It is here noted that among the possible infinite choices to construct hybrid models, we selected those that, in our best judgment, reflect attractive possibilities to utilize available knowledge and commonly measured variables in the context of cell cultures.

Thus, when moving from 0 to 100% degree of hybridization, the number of underlying phenomena explicitly described in the hybrid model increases such that the most general assumptions and phenomena are described first. Subsequently, the purely DD model (0% hybridized) has no phenomena incorporated and, thus, tries to predict the process variables directly. On the other hand, the "Hybrid-No Mass Balance (Hyb-No MB)" model adds the first fragment of knowledge by modeling the rate of accumulation using the DD model. Thus, the neural network (NN) learns the entire right-hand side of the rate equations, represented by a system of ordinary differential equations (ODEs). In other words, the NN must learn to close the mass balance (or conserve the mass) and is not imposed a priori (thus the name No Mass Balance).

However, the next hybridized model, the "Hybrid (Hyb)" model, explicitly states the mass balance by taking into account the inflow and outflow of different process variables from the bioreactor. Thus, it only learns the rate of consumption (or production) of different process variables in the system using a NN instead of the entire mass balance learned by the Hyb-No MB model. Adding a further piece of knowledge, the "Hybrid Rate (HR)" model encodes the proportionality of the reaction rates to X_v. Therefore, it learns only the cell-specific rates of consumption (or production) of cells, metabolites, and products. Furthermore, it is known that cells grow and subsequently die. Thus, the cell-specific rate for the viable cell density can further be learned as a specific rate of growth and death of cells. This is incorporated in the "Hybrid X_v Decomposition (HXD)" model, in which the NN learns the specific growth and death rates separately (NN_{1a}, NN_{1b} in Table 1).

The HR and HXD models are developed on the limiting assumption of only proportionality with the viable cell density for metabolites while it is known that it is possible to have kinetic consumption (or production) of metabolites via pathways that are not dependent on viable cell density[17,18] (e.g., direct degradation of glutamine to ammonia).[62] Thus, the "hybrid all decomposition (HAD)" model incorporates further knowledge of having cell-independent kinetic pathways for consumption/production of metabolites, thus having an increased degree of hybridization. The HAD model subsequently allows for two contributions to the reaction production or consumption term for the metabolites are considered: (i) the consumption or production directly related to the viable cells and (ii) the consumption or production through other pathways, not directly related to the viable cell density. Finally, on the right extreme (100% hybridized) is the mechanistic model (Mech), which replaces the NN in the HAD model with empirical equations[18,62,63] such as Monod

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

type kinetics with inhibition are used for the specific cell growth and death rates.

We note here that different statistical methods can be used within the hybrid model framework, thus giving rise to a parallel family along the "method axis". This concept is already well established in the existing literature[27,28,31,32,64,65] and is out of the scope of this work. Consequently, in the following, we use only one common type of statistical method, namely, neural networks (NN), as the data-driven part within the hybrid models. Additionally, general modeling assumptions such as ideally mixed liquid phase, homogeneous cell population, and accurate feeding information assumptions apply to all of the models.

The following section briefly summarizes the training procedure of the different models. All computations are performed in MATLAB R2020b. It is noted that titer was not used among the input variables to model other process variables or themselves. First, it is of interest to be able to use these models as a soft sensor for titer to replace the analytical quantification in the future. Second, titer, which is the product secreted by the cells, is expected to have no impact on other process variables (e.g., metabolite consumption/production and cell growth).

2.2.1. Data-Driven Models. The batch-wise unfold-partial least squares 1 (BWU-PLS1) model is considered representative of the data-driven (DD) approach. This choice is made on a pre-analysis, where four different DD models: black box-PLS1 (BB-PLS1), BWU-PLS1, artificial neural network (ANN), and variable wise unfold-artificial neural network (VWU-ANN), were compared based on their predictive performance on training sets of different sizes. It is here noted that approaches based on only PLS1 and ANN were tested for DD modeling because PLS1 is still the algorithm predominantly used for process modeling in the industry for cell cultures, while bioprocess modeling with ANNs has been reported in the literature as a benchmark to compare the hybrid model performance.[40] Additionally, checking the performance of ANNs was reasonable, given ANNs were used to model the DD part in the hybrid models. The details of the model equations and the performance of the different DD models are provided in the Supporting Information (SI) (Sections S1 and S2).

In BWU-PLS1, process data are unfolded batch-wise, i.e., each row is representative of a single experiment, and columns comprise different process variables measured at different times.

For each process variable at each time point, a PLS1 model is trained using the previously obtained BWU matrix embedding historical information available until that given time point. A detailed discussion of the BWU-PLS1 models can be found in refs 54 and 59 and has been explained in Section S1.2, Supporting Information. The optimal number of latent variables for each PLS1 model is chosen based on four-fold cross-validation.

2.2.2. Hybrid Models with Intermediate Degree of Hybridization. The different hybrid models are set up as indicated in Table 1—column 3, with the NN receiving the inputs as summarized in Table 1—column 4. The inputs are scaled to variable-specific scaling factors to ensure numerical stability in optimization. A multioutput, single-layer feedforward NN is used in this work with the number of nodes as the hyperparameter. The optimal hyperparameter is chosen for each hybrid model variant based on the inspection of the performance of the models on a validation set. For the in silico case study, a single validation set sampled using an LHC design is used. For the experimental dataset, a four-fold cross-validation approach is adopted. Each node in the hidden layer presents a tanh() activation function, with the final output layer performing a linear transformation. For HXD and HAD models (which learn the specific growth and death rates of cells separately), a ReLU activation is applied in the output layer to ensure negative growth rates and positive death rates are not learned. The hybrid models are solved using a nonlinear parameter estimation problem using fminunc() based on a quasi-Newton optimization algorithm.

A further detailed description of the training procedure for developing the hybrid models can be found in our previous work.[9,42]

2.2.3. Mechanistic Models. To obtain the parameters in the mechanistic model, nonlinear parameter optimization is performed using the in-built fmincon() function, which is based on the interior-point algorithm.

2.3. Model Performance Metric. The models are compared based on an overall scaled mean squared error obtained across all of the runs (N_runs), times (N_time), and variables (N_var), as indicated below

$$\mathrm{MSEP}_{\mathrm{overall}} = \frac{1}{(N\_runs \cdot N\_time \cdot N\_var)} \sum_{n=1}^{N\_runs} \sum_{t=1}^{N\_time} \sum_{i=1}^{N\_var} \left(\frac{C_{i,t,n}^{\mathrm{meas}} - C_{i,t,n}^{\mathrm{pred}}}{\mathrm{scaling}_i} \right)^2 \quad (1)$$

where $C_{i,t,n}^{\mathrm{meas}}$ and $C_{i,t,n}^{\mathrm{pred}}$ are the in silico measurement and the model prediction of the ith variable for tth time and nth run, respectively. The variables predicted using the different models, for both the in silico and experimental datasets, are summarized in Figure 1A. $\mathrm{Scaling}_i$ is a scaling factor for the ith variable used to normalize the value of the different process variables. Since the process variables have different magnitudes, scaling the contribution to the error coming from each process variable is important.

For the in silico dataset, both the true (i.e., simulated) and the measured (i.e., perturbed with noise) profiles are available. Therefore, we can compute the $\mathrm{MSEP}_{\mathrm{overall}}$ metric for the analytical (or measurement) error, $\mathrm{MSEP}_{\mathrm{overall,\ analytical}}$, that would indicate the predictive limit that can be achieved by any model. $\mathrm{MSEP}_{\mathrm{overall,\ analytical}}$ can be computed as follows

$$\mathrm{MSEP}_{\mathrm{overall,analytical}} = \frac{1}{(N\_runs \cdot N\_time \cdot N\_var)} \sum_{n=1}^{N\_runs} \sum_{t=1}^{N\_time} \sum_{i=1}^{N\_var} \left(\frac{C_{i,t,n}^{\mathrm{meas}} - C_{i,t,n}^{\mathrm{true}}}{\mathrm{scaling}_i} \right)^2 \quad (2)$$

3. RESULTS AND DISCUSSION

3.1. Model Accuracy and Cost along the Hybridization Axis: In Silico Data. First, we characterize the predictive capability of the different models and their "cost", that is, the minimum number of runs required to build an optimal version of each of them. The computational resources (time, number of CPUs, etc.) to develop a model depend on the size of the training data. In addition, the ability to generate data is often the bottleneck. Thus, we here choose the number of experiments in the training set as the representative metric

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

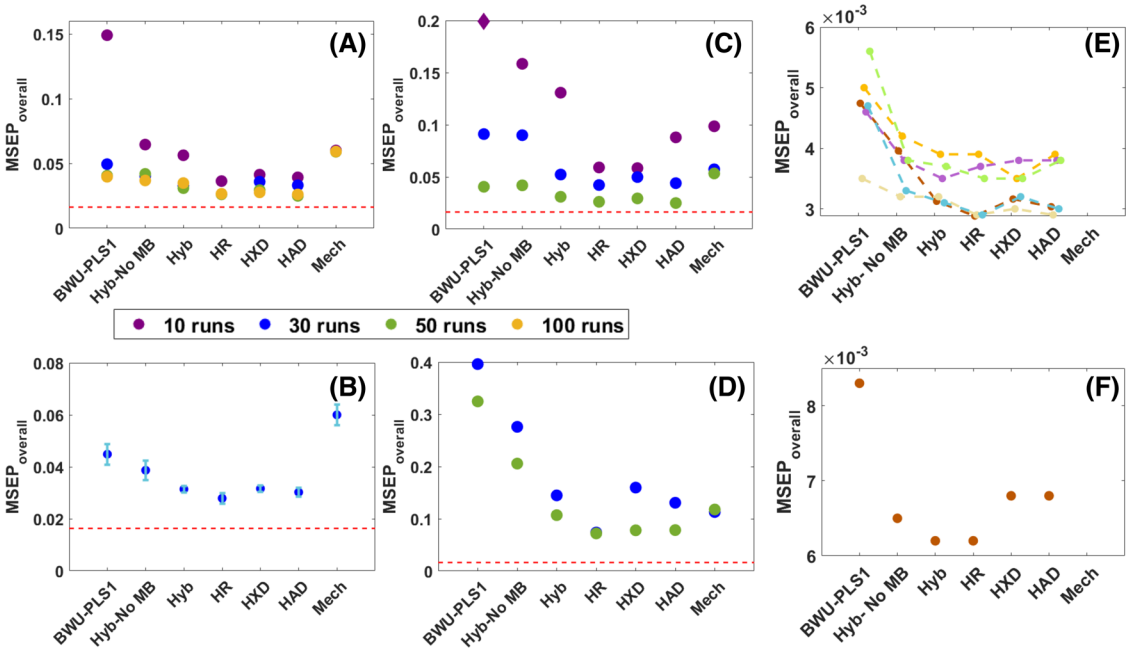


Figure 2. Comparison of the different models, namely, BWU-PLS1 model, various hybrid models, and Mech model in terms of (A) accuracy in predicting the interpolation test set using models trained with different numbers of runs: 10, 30, 50, and 100; (B) robustness in interpolation prediction illustrated by creating five realizations of the training dataset and testing on the interpolation test set. (C) Extrapolation capability along the initial condition of X_v, GLC, and GLN and (D) extrapolation capability along the feed bolus amount and feeding days. (E) Interpolation accuracy on an experimental dataset, where the different lines indicate the performance of the models trained on five different training sets and subsequently tested on five different test sets. (F) Extrapolation capability illustrated on an experimental dataset.

for model building cost. For this purpose, training datasets of sizes 10, 20, 30, 50, 70, 90, and 100 runs were tested. In Figure 2A, the predictive accuracy of the models with different degrees of hybridization (x-axis) are compared, based on the $MSEP_{overall}$ metric (y-axis), for different numbers of runs in the training set: 10, 30, 50, and 100. Models are tested on the interpolation test set (cf. Figure 1B).

It is here noted that due to the computational time required for training the different hybrid models (~hours), performing multiple repetitions was not feasible. However, for one exemplary case (30 experiments), the selection of training set (using the LHC design) was repeated five different times, with different random seeds. The predictive accuracy on the interpolation test set of the different models along the hybridization axis is represented in Figure 2B. It can be observed that all of the models have a smaller standard deviation indicating the models reached a similar performance. This is likely because a space-filling design is used to create the training set, which samples the design space to perturb similar interactions whenever generating the data.

3.1.1. Low Data: Data-Driven and Hybrid Models. In Figure 2A, it is seen that when only 10 runs are available for training the model (violet circles), the predictive performance of the BWU-PLS1 model is poor, with an $MSEP_{overall}$ of about 0.150. However, with increasing amounts of process knowledge, from Hyb-No MB to Hyb to HR, the predictive performance of the model improves considerably. A similar improvement to the HR model is observed for the two hybrid models with the largest degree of hybridization, the models HXD and HAD. In particular, when trained with 10 runs, the HR, HXD, and HAD models exhibit an $MSEP_{overall}$ = 0.036, 0.039, and 0.040, respectively, which are comparable to the best predictive performance achieved by the BWU-PLS1 model obtained with a training data set of 50 runs. This highlights the substantial improvement in the predictability of hybrid models compared to purely data-driven models and confirms that low availability of data, or low cost of model development, can be compensated by incorporating process knowledge to improve predictive performance.

3.1.2. Mechanistic Model Behavior. When we analyze the predictive performance of the purely mechanistic model (Mech in Figure 2A), we first notice that the performance is invariant to the number of runs. This can be explained by the requirement of estimating only a few model parameters in a purely mechanistic model. Thus, even in the presence of a limited amount of data (e.g., 10 runs), the data is always in excess with respect to the parameters, making the estimation problem mathematically well defined. However, the performance of the mechanistic model, which shows an $MSEP_{overall}$ of about 0.060, is significantly worse compared to the best hybrid models, although still better than the purely data-driven model with only a few (10) runs. This can be explained in terms of model bias. In other words, the assumptions underlying the model, such as the use of simple Monod equations and constant kinetic parameters, is not fully representative of the actual mechanism of the process, thus leading to poor performance. Mechanistic models could be trained with very little data and, thus, in theory, would have a low cost if an unbiased model is already available. This is rarely the case for such complex and poorly understood processes. Therefore, even relatively complex mechanistic models are likely to contain some biases that can possibly be corrected only with some further costly and sophisticated experimental investigation.

3.1.3. Minimum Training Data Required to Reach Analytical Limits. As mentioned in Section 2.3, the $MSEP_{overall}$

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

metric for the analytical measurement error is $MSEP_{overall,analytical}$, which is 0.020 for the considered dataset (red dashed line, Figure 2A). In other words, this can be considered as the predictive limit for any model and is thus used to benchmark the predictive performance of the different models. The BWU-PLS1 model required 50 runs to achieve its best predictive capability, after which the model had a constant prediction metric $MSEP_{overall}$ = 0.039. This is comparable to the performance of the simplest hybrid model (Hyb-No MB) with 30 runs and that of the best hybrid model (HR) with only 10 runs. The fact that the BWU-PLS1 model cannot further improve, irrespectively of the number of runs added beyond 50, can be explained by the intrinsic linear nature of the model, which is biasing the model and limiting its performance. When the "mechanistic" description of the feeding process is explicitly incorporated in the Hyb model as a mass balance, the $MSEP_{overall}$ reduces to 0.03, and it requires only 30 runs to achieve its best predictive capability. Proceeding in this direction, we observe that the HR, HXD, and HAD models reach their best performance with an $MSEP_{overall}$ of 0.025, which is relatively close to the measurement error of the analytics. The analysis of Figure 2A provided a quantitative demonstration of the predictive capability of hybrid models. The increase in the degree of hybridization is generally beneficial in reducing the cost of the model, as long as the introduced process knowledge is not such that a strong bias (or excessive constraining of the mathematical structure) is introduced in the model as in the Mech model.

It should also be noted that the HR hybrid model required only 30 runs to achieve its best predictive capability, while the HXD and HAD models required 50 runs. The addition of further process knowledge in the HXD and HAD models comes at the cost of increased network size (i.e., more outputs) and thus more parameters. This introduces additional flexibility in the models, which requires more runs (50 instead of 30) to learn a generalizable behavior. The underlying NN in the HR model is flexible enough to explain very complex relationships. Thus, adding two NNs is not likely to add any improvement in predictive accuracy in the training region (interpolation), but just an additional experimental cost.

Overall, at least for the cell culture studies considered in this work, we can conclude that, among the different options available, the HR model presents an optimal trade-off between model accuracy and cost in terms of incorporating process knowledge.

3.2. Extrapolation Performance along the Hybridization Axis: In Silico Data. Following the comparison of the predictive capabilities in the interpolation of the different models, the predictive performance in extrapolation is assessed, which is crucial for process design and optimization applications, where the models may be used to predict operating conditions in the design space not explored during training.

3.2.1. Extrapolation along the Initial Condition. As a first comparison, we consider a test set where the initial conditions of X_v, GLC, and GLN are outside the design space used for training the models. This is shown in Figure 2C, where the $MSEP_{overall}$ metric evaluated on the extrapolation 1 test set (cf. Figure 1B) is compared for models with different degrees of hybridization. Three scenarios are highlighted: performance of models trained with (i) small training data set: 10 runs, (ii) reasonable training data set: 30 runs, and (iii) size of training data set for best interpolation accuracy: 50 runs. Since the model already achieved the best predictive accuracy using 50 runs, the study of models trained with 100 runs is not presented (to avoid overcrowding of the plots).

3.2.1.1. Small Training Data Set. We start by comparing the performance of the different models in the case of small data sets (10 runs, violet symbols). The poor performance of BWU-PLS1 observed in the case of interpolation (Figure 2A) for small training data set is confirmed and amplified in the case of extrapolation, resulting in a very high $MSEP_{overall}$ of 0.700, as indicated by the purple diamond in Figure 2C (out of scale). When transiting from BWU-PLS1 to the HR model through the Hyb-No MB and the Hyb models, the extrapolation performance follows a similar trend to that of the interpolation case. The $MSEP_{overall}$ reduces with increasing addition of process knowledge, with the HR model resulting in an $MSEP_{overall}$ of 0.060. When further increasing the degree of hybridization, this trend changes: the HXD model shows an extrapolation accuracy similar to the HR model, while the $MSEP_{overall}$ of the HAD model increases to about 0.08. This can be attributed to the fact that the HAD model has more parameters, and therefore training with only 10 runs might lead to overfitting of the training design space resulting in a relative deterioration of the prediction performance in extrapolation. Finally, the Mech model at the right end of the hybridization axis shows an $MSEP_{overall}$ of 0.09, which is slightly larger than the HAD model but 1.5 times larger than other hybrid models such as the HR and HXD models.

3.2.1.2. Sufficient Training Data Set. On the other hand, when a very large training data set is considered so as to enable the training of robust models (50 runs, green symbols in Figure 2C), the hybrid models with mass balances (Hyb, HR, HXD, and HAD) show a considerably better extrapolation accuracy, resulting in an $MSEP_{overall}$ of 0.025−0.030, which is approaching the benchmark set by the analytical error. In contrast, the BWU-PLS1 and Hyb-No MB models show an $MSEP_{overall}$ of about 0.040, while the Mech model exhibits an $MSEP_{overall}$ of 0.060. It has to be noted that, when extrapolating along the initial condition of X_v, GLC, and GLN, the different models achieve an extrapolation prediction accuracy comparable to their respective interpolation prediction accuracy, provided that the training data set is sufficiently large.

3.2.1.3. Reasonable Training Data Set. However, when considering the typical size of the training data sets available for cell culture modeling (e.g., 30 runs and blue symbols in Figure 2C), it is seen that both the BWU-PLS1 and Hyb-No MB model exhibit poor extrapolation performance leading to $MSEP_{overall}$ of about 0.09. Adding mass balance in the Hyb model significantly improves the extrapolation performance of the model resulting in an $MSEP_{overall}$ of 0.050. A further increase in the degree of hybridization improves the hybrid model performance, with the HR model having an $MSEP_{overall}$ of 0.040. However, beyond this, the trend changes again, with the HXD and HAD models exhibiting a slight increase in the $MSEP_{overall}$ to 0.045. This is due to the additional flexibility introduced in the model through a larger number of parameters, which requires more training data to learn a generalizable behavior. Finally, the Mech model that marks the right extreme of the hybridization axis has an even larger error in extrapolation with $MSEP_{overall}$ = 0.060.

3.2.2. Extrapolation along the Feeding Strategy. A similar behavior, but even more pronounced, is observed when analyzing the extrapolation along the feeding conditions (extrapolation 2 dataset, Figure 1B). This includes the amount

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

of bolus feed on the different days of the culture and the first and last day of feeding. Figure 2D shows the $MSEP_{overall}$ of the different models trained with a reasonable (30 runs, blue symbols) and sufficient (50 runs, green symbols) number of runs. Since a more pronounced deterioration of model performance was observed when extrapolating along the feeding strategy, the models trained with 10 experiments (that already showed poor accuracy for a simple extrapolation for the initial conditions) are left out.

Even with 50 training runs, a purely data-driven model, such as the BWU-PLS, is incapable of handling the feed extrapolation, resulting in a very large $MSEP_{overall}$ of 0.320. However, when modeling the dynamic change in the concentrations of the process variable, such as in the Hyb-No MB, the performance improves significantly, leading to a lower $MSEP_{overall}$ of about 0.200, which is however still too large for practical applications. Moreover, for models trained with only 30 runs, the performances of both the BWU-PLS1 and the Hyb-No MB models are poor. The change in the feeding strategy is likely to affect the correlation structures learned by the PLS1 in the DD model and by the NN in the hybrid model, which potentially results in correlations among variables that are not identified during model training (with either 30 or 50 runs).

In contrast, the Hyb model showed an $MSEP_{overall}$ of 0.100 in feed extrapolation when trained with 50 experiments (and a slightly higher $MSEP_{overall}$ of 0.120, when trained with 30 experiments), highlighting the importance of encoding information in a physically relevant manner, as discussed in detail in Section 3.4. Adding further physical constraints improves the model performance in feed extrapolation, resulting in an $MSEP_{overall}$ of 0.050, as demonstrated by the HR, HXD, and HAD models. However, limited by the strict functional imposition in the Mech model, the $MSEP_{overall}$ again raises to 0.100. While the HR model has a comparable performance when trained with 30 or 50 runs, the HXD and HAD models show a larger $MSEP_{overall}$ of 0.100 when trained with only 30 runs. This is again due to the requirement of more data to address the additional flexibility introduced by adding more process knowledge in this model. It is worth noting that, in the case of feed extrapolation, the models do not achieve the same level of accuracy as in interpolation prediction. This is because, while extrapolating the feeding strategy (mainly due to the change in the feed starting and ending times), it is likely that the model encounters a combination of process variables that were not observed during training.

Thus summarizing, with the typical sizes of the training sets, both the left and right ends of the hybridization axis show poor performance. Adding process knowledge (starting from no knowledge in a DD model) improves the model extrapolation prediction accuracy, in particular, hybrid models with mass balances show better extrapolation accuracy, but this goes through a maximum as the hybridization degree increases further. Among all different hybrid models, the HR model shows the best performance in extrapolation prediction both across the initial condition and feeding strategy. For the considered system, this corresponds to the best degree of hybridization, similar to the observation made earlier in the case of interpolation predictions.

3.3. Experimental Case Study of Industrial Relevance. In this section, we now analyze the interpolation and extrapolation behavior of hybrid models with different degrees of hybridization with respect to an industrial lab-scale bioreactor experimental dataset. This dataset consists of designed variables, such as the $N - 1$ parameters, DO, pH, and process variables, such as osmolality, which cannot be predicted with the mechanistic model (Mech) used in this work. Therefore, the Mech model is not included in the following analysis.

Figure 2E compares the interpolation capabilities of the different models along the hybridization axis. It is here noted that five different train-test splits were performed using the approach mentioned in Section 2.2.2. Subsequently, all of the models were trained on the five training sets and tested on the respective test set. The different colored lines in Figure 2E indicate the $MSEP_{overall}$ trends for these five test sets. A qualitative behavior similar to that found for the in silico data set is observed, with the BWU-PLS1 model having the highest $MSEP_{overall}$. This decreases with the addition of process knowledge until the HR model, which exhibits the lowest $MSEP_{overall}$. Beyond this, the HXD and HAD models have a comparable or slightly higher $MSEP_{overall}$ to that of the HR model (cf. Figure 2E). It is to be noted that only in one out of the five cases, the BWU-PLS1 model showed interpolation performance similar to the hybrid models (beige line in Figure 2E). For all of the other cases, it shows a significantly higher $MSEP_{overall}$.

As a next step, the extrapolation performance of the different models along the hybridization axis is assessed on the extrapolation test set (cf. Figure 1B,C), and the results are summarized in Figure 2F. Considering the intrinsic difficulty of extrapolating across different process behavior (excluding the five yellow profiles), we expected to observe a general deterioration of the model performances compared to the interpolation case. Despite this, the BWU-PLS1 model still exhibits the worst performance with an $MSEP_{overall}$ of 8.3×10^{-3}. The addition of process knowledge improves the model performance by about 20%, reducing the $MSEP_{overall}$ to about 6.5×10^{-3} for the Hyb-No MB model and further down to 6.2×10^{-3} for the Hyb and HR models. However, for the HAD and HXD, the extrapolation error is slightly higher compared to the Hyb-No MB, Hyb, and the HR model resulting in an $MSEP_{overall}$ of about 6.8×10^{-3}. This is expected, given the complexity of the HXD and HAD model is higher and therefore requires a higher amount of training runs to obtain a generalizable model. With the current training size of 26 experiments, with only five experiments being similar to the predicted experiments, the model likely encountered some overfitting issues (despite regularization), resulting in a lower extrapolation accuracy. In spite of this, the example confirms the importance of using hybrid models compared to purely data-driven approaches for real data, especially for process development and optimization.

3.4. Comparison of Extrapolation Capability among Hybrid Models. From the two extrapolation case studies discussed above, we can conclude that two critical phenomena or modeling assumptions need to be described explicitly in mechanistic form when developing hybrid models for cell cultures. The first one corresponds to the incorporation of the feeding procedure in the form of mass balances and the second is the incorporation of the proportionality of the metabolic rates with respect to the viable cell density, X_v. Such a formulation allows the study of cell-specific rates, which is crucial for understanding the metabolic consumption (or production) tendencies of the cells. To further demonstrate

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

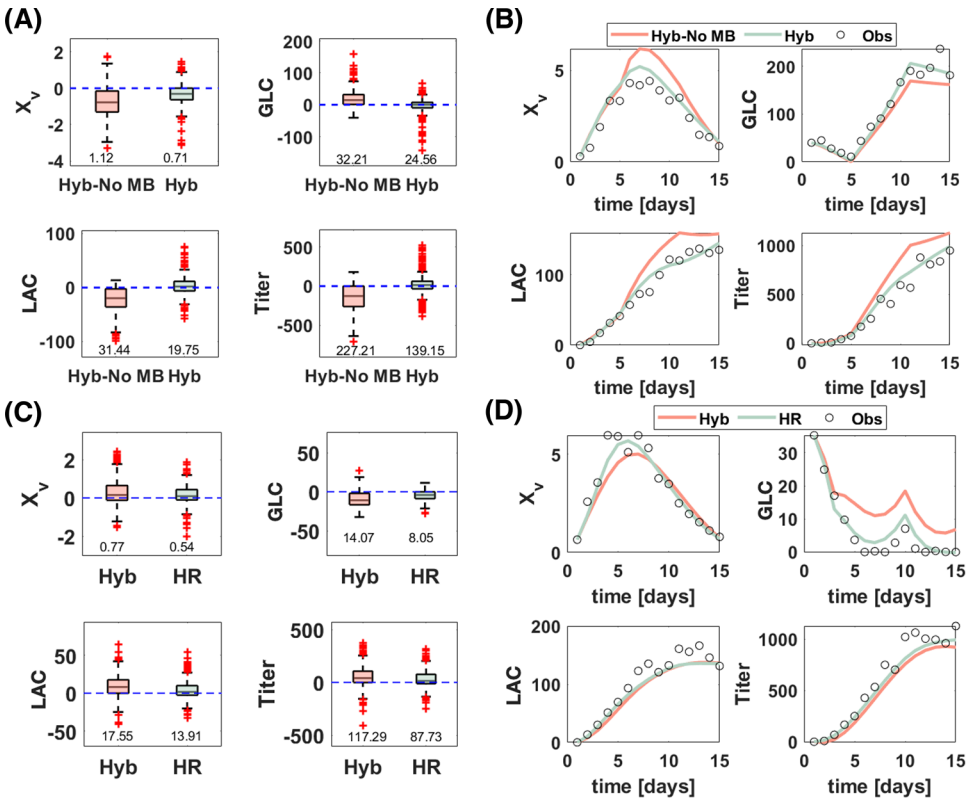


Figure 3. Comparison among hybrid models for four key variables, namely, X_v, GLC, LAC, and Titer: comparison between the Hyb-No MB and Hyb models illustrated through (A) boxplots of residuals and quantifies the RMSEP, and (B) the dynamic evolution of an exemplary run. Comparison between the Hyb and HR model illustrated through (C) boxplots of the residuals and quantifies the RMSEP, and (D) the dynamic evolution of an exemplary run.

the implication of these two fragments of knowledge (or modeling assumption), we compare the performance of selected hybrid models on specific extrapolation test sets (extrapolation 2 and extrapolation 3 test sets). Boxplot of residuals (residual = obs − pred) and root mean squared error in prediction (RMSEP) for four key process variables: X_v, GLC, LAC, and Titer, are used to compare the models. The bias can be analyzed by observing the distance of the median from zero, while the variance can be interpreted through the spread of the distribution.

For the first case, we compare Hyb-No MB and Hyb model on a feed extrapolation test set with 50 runs (extrapolation 2 test set). This is shown in Figure 3A, where the boxplots of residuals are compared for the Hyb-No MB (red) and Hyb (green) models. It can be observed that the Hyb-No MB model has a higher RMSEP, bias, and variance for all of the four process variables in comparison to the Hyb model. Further, the distribution for X_v, LAC, and Titer for the Hyb-No MB model lies below the zero line, indicating that the values predicted are higher than the observed values. Thus, the model overpredicts these variables while following the same reasoning it underpredicts GLC. In contrast, the Hyb model does not show any significant bias for any of the process variables except for X_v, where the Hyb model slightly overpredicts. For all of the process variables, the Hyb model shows a much more narrow distribution of residuals compared to those of the Hyb-No MB model. These behaviors are confirmed by the time evolutions predicted by the two models, as shown in Figure 3B using an exemplary run. The Hyb model, in general, shows very good agreement with the observations, in contrast to the Hyb-No MB model, which instead exhibits significant deviations. Despite the apparent simplicity of the added knowledge, we can then conclude that the incorporation of the feeding mechanism in the form of a mass balance is crucial for the model performance, as shown here for feed extrapolation. This is relevant for transferring the models across different feeding operating modes, as discussed in detail in Section 3.5.

To demonstrate the advantage of introducing cell-specific rates, we compare the performance of Hyb and HR models on the extrapolation 3 test set, where the initial condition of X_v is sampled outside the training region. Subsequently, the comparison of the Hyb and HR model is shown in Figure 3C again based on the boxplot of residuals and RMSEP. It can be observed that, between the two, the Hyb model has a larger RMSEP, bias, and variance for all of the four considered process variables though the bias and variance are smaller compared to the previous case. Following the same strategy for analysis, the Hyb model overpredicts GLC and underpredicts LAC and titer. These observations are also reflected in the dynamic predictions of the Hyb model compared to the HR model, as shown for an exemplary run in Figure 3D. Thus, when not using cell-specific rates, the NN in the framework must learn the complex interactions between the cell-specific rates and the viable cell density purely from the data. Consequently, when extrapolating, the NN in the Hyb model faces unseen interactions among process variables, thus highlighting the importance of separating primary effects (of variables) whenever this is possible.

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

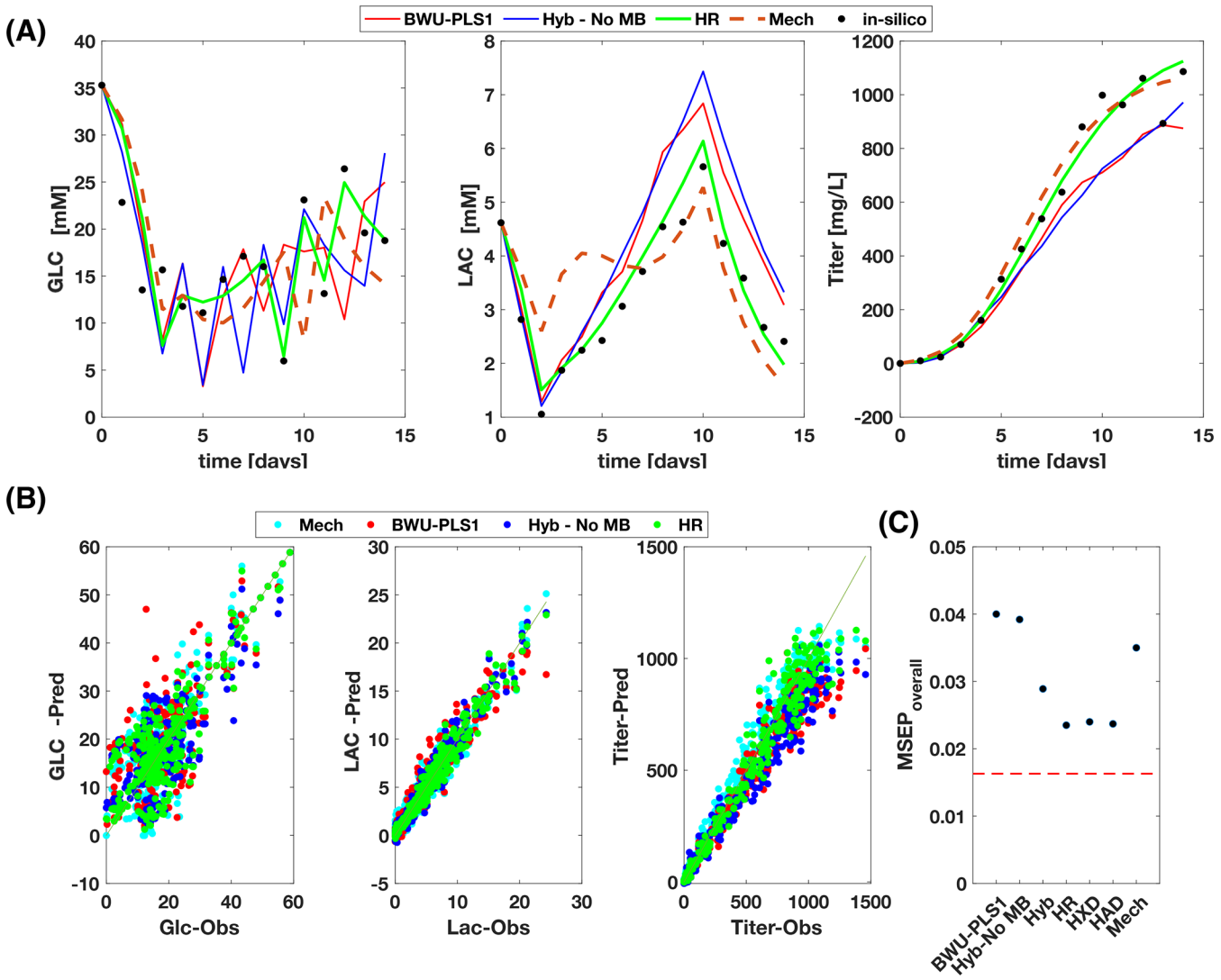


Figure 4. Use of the different models trained in pulsed fed-batch mode, for conditional GLC feeding strategy. (A) Comparison of the dynamic evolution of GLC, LAC, and Titer resulting from the feeding strategy defined by the BWU-PLS1, Hyb-No MB, HR, and Mech models. (B) Parity plots of GLC, LAC, and Titer comparing the above results relative to the BWU-PLS1, Hyb-No MB, HR, and Mech models with the in silico model results. (C) $MSEP_{overall}$ values based on the predictions given by the different models ordered by increasing degree of hybridization.

3.5. Application: Fixed Bolus Feeding to Conditional Feeding in Fed-Batch. The motivation behind the discussion of this study case is to show that the predictive and extrapolation capabilities of a model have a critical effect in applications requiring trustworthy modeling solutions to support near-real-time decision making, process knowledge transfer, control, and optimization. In particular, incorporating the feeding strategy in the model through the mass balances is extremely important to facilitate the transfer of knowledge across bioreactor operation modes, efficient process control (GLC being an important controlled parameter), and process optimization (i.e., feeding strategy).

As a result, the models trained in the frame of a pulsed fed-batch operating mode can be transferred to a continuous fed-batch or even a perfusion operating mode or feeding strategies with no fixed time structure, simply through an obvious reformulation of the feeding terms in the relevant mass balances. In this work, this characteristic is offered by the hybrid models, namely, Hyb, HR, HXD, and HAD. It is here to be noted that the cell-related specificities that arise with the change in the mode of operation (e.g., cell aging and mutation in perfusion) might have to be additionally learned for the new mode of operation.

This ability of model transferability is illustrated using a modest case study of applying a conditional daily bolus feeding to the process. According to this strategy, GLC is measured at the end of every day and, if it is less than a threshold (in our example, 15 mM), a suitable GLC pulse is given to the bioreactor in the form of bolus feed to reach a second threshold concentration (30 mM in our example). The presence of such conditionality makes the experiment evolution different from the unconditional feed strategy used for model training and thus difficult to predict.

Figure 4A demonstrates a comparison of the GLC, LAC, and Titer evolution, obtained using BWU-PLS1, Hyb-No MB, HR, and Mech models to implement a conditional feeding strategy. This is compared to the simulated profile obtained from the in silico model under the same conditions (black symbol). It is seen that, for this exemplary experiment, the dynamic evolution for GLC predicted by the BWU-PLS1, Hyb-No MB, and Mech models is unable to match the evolution of the in silico model. Of course, this affects the

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

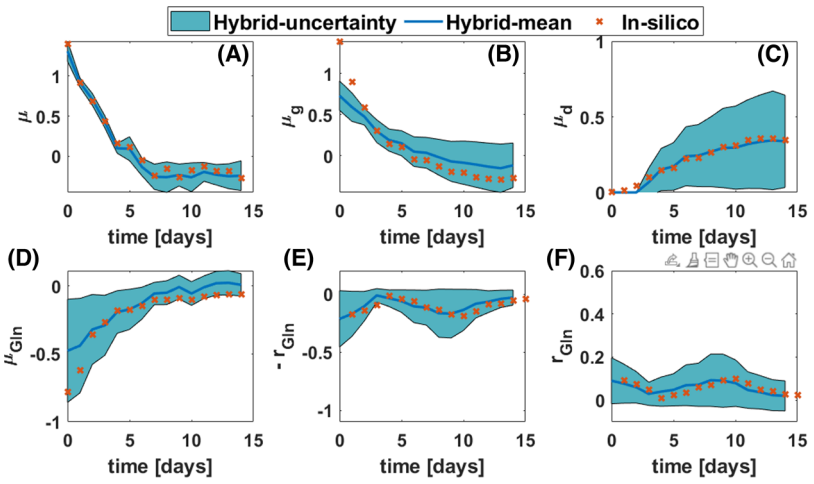


Figure 5. Prediction of metabolic rates using different hybrid models (blue) compared with the in silico model results (red). The blue line indicates the mean prediction from five different realizations of the respective hybrid models, and the shaded region represents twice the standard deviation made in the prediction by the five different realizations: (A) the lumped specific rate of cell production (μ) using the HR model, (B) the specific growth rate (μ_g) of viable cells using HXD model, (C) the specific death rate (μ_d) of viable cells using HXD model, (D) the specific rate of consumption of glutamine for cell growth (μ_{gln}) using the HAD model, (E) the additional consumption of glutamine for NH_4 production ($-r_{GLN}$) using the HAD model, and (F) the production of NH_4 from direct GLN consumption (r_{GLN}) using the HAD model.

prediction of variables that are strongly dependent on the GLC concentration, such as LAC and Titer. Therefore, these models may not be suitable for practical applications involving changes in feeding strategy or operation modes. On the other hand, the dynamic evolution predicted by the HR model (green line) is coherent with that of the in silico model due to its superior predictive and extrapolation capabilities. Figure 4B presents the observed (i.e., predicted by the in silico model) vs predicted plots of the model predictions for GLC, LAC, and Titer. It can be observed that both BWU-PLS1 and Hyb-No MB underpredict titer while the Mech model overpredicts it, thus showing a systematic bias. For GLC and LAC, the BWU-PLS1 and Hyb-No MB have a broader variance followed by the Mech model. The HR model, however, shows no bias for either case and has the least variance in comparison to all of the others. Note that since the HXD and HAD models perform similarly to the HR models, their predictions are not shown to avoid visual confusion. However, Figure 4C summarizes the performance of all models, including these two, in terms of the $MSEP_{overall}$ as a function of the hybridization degree.

3.6. Improving Process Understanding. Another factor that needs to be considered about hybrid models is their potential to generate new knowledge or a better understanding of the process. In particular, the HR, HXD, and HAD models provide the opportunity to investigate the effect of changes in the cell culture operating conditions and process variables on macroscopic metabolic properties such as the specific rates of particular metabolic processes. The QbD initiative emphasizes the requirement for a better understanding of the effect of process parameters on quality attributes. Thus, the ability to extract such information is critical toward understanding the variable interrelationships and dynamics of a process better.

In this regard, Figure 5 compares the macroscopic information extracted from different hybrid models with the corresponding values obtained from the in silico model used to generate data in this work. In particular, the following equations in the in silico model are relevant for the comparison

$$\frac{dX_v}{dt} = \mu X_v, \text{ such that } \mu = \mu_g - \mu_d \tag{3}$$

$$\frac{dX_v}{dt} = \mu_g X_v - \mu_d X_v \tag{4}$$

$$\frac{dGLN}{dt} = \mu_{gln} X_v - r_{GLN} \tag{5}$$

$$\frac{dNH_4}{dt} = \mu_{NH_4} X_v + r_{GLN} \tag{6}$$

where μ is the lumped specific rate of cell production (growth and death), μ_g is the specific growth rate, μ_d is the specific death rate, μ_{gln} represents the specific production rate of glutamine, μ_{NH_4} is the specific production rate of NH_4, and r_{GLN} represents the cell-independent kinetic degradation of glutamine into ammonia as proposed in refs 18 and 62. In the in silico model, each of these terms, μ_g, μ_d, μ_{gln}, μ_{NH_4}, and r_{GLN}, are described through empirical expression as a function of X, Z, and W information variables. For an exemplary run in the test set, the system of ODE is solved and the corresponding value of the μ, μ_g, μ_d, μ_{gln}, μ_{NH_4}, and r_{GLN} is computed for the combination of process variables at each time step using these empirical expressions. Such an approach is chosen in contrast to computing the values for random combinations of process variables, to have physically relevant combinations of different process variables. Following a similar strategy, the relevant outputs of the NNs from the different hybrid models are extracted while solving the system of ODE for the same test run. Hereby, five different training sets are selected by repeating the LHC sampling five times with different random seeds and subsequently five hybrid models are trained. Thus, Figure 5 indicates the mean and the standard deviation of the relevant macroscopic parameters extracted from these five repetitions.

In Figure 5A, the prediction of the lumped specific rate of cell production (μ) predicted by the HR model (Table 1, row 4) is compared to the actual values from the in silico model.

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

Table 2. Performances of the Models Developed in This Work, with Increasing Degree of Hybridization, with Respect to Different Applications[a]

	Model ranking								
	Low data (10 runs)		Reasonable data (30 runs)						
Model	Inter	Extra	Inter	Extra	Reach analytical limit (50 runs)	Min no. runs	Interpretation	Model transfer (operation mode)	Model transfer (product/process)
BWU-PLS1	5	6	4	5	3	50		no	No
Hyb-No MB	4	5	3	5	3	30		no	Yes
Hyb	3	4	2	3	2	30	reaction rate	yes	Yes
HR	1	1	1	1	1	30	μ_i	yes	Yes
HXD	2	1	2	2	1	50	μ_i, μ_g, μ_d	yes	Yes
HAD	2	2	2	2	1	50	μ_i, μ_g, μ_d, r_i	yes	yes**
Mech	4	3	5	4	4	10	kinetic constants and physicochemical parameters	yes**	

[a]The column interpretation reports the process understanding learned by the model in addition to the knowledge imposed explicitly during the setup of the hybrid model. "Inter" is used to refer to interpolation predictions, while "extra" refers to extrapolation predictions. Columns "low data" and "reasonable data" refer to the training data sets; in model ranking, "1" indicates the best performing model among the six considered, while "6" indicates the worst performing model. μ_i is the cell-specific rate for the different metabolites and viable cells, μ_g is the specific cell growth rate, μ_d is the specific cell death rate, and r_i is the kinetic rate of consumption/production of metabolites independent of cell growth. Note: ** indicates that the outcome is theoretically possible but is subject to the accuracy of the model.

For this, the output of the NN corresponding to the viable cell density (cf. Table 1) NN_{X_v} is extracted and compared to the μ used in the in silico model (eq 3). Next, in Figure 5B,C, a similar comparison is illustrated but using the HXD model predictions, which independently model the specific cell growth (NN_{1a}) and death (NN_{1b}) rates. These values are, respectively, compared to the specific growth (μ_g) and death (μ_d) rate of cells obtained from the in silico model (cf. eq 4). Finally, in Figure 5D−F, the following quantities are compared: (i) The specific consumption rate of glutamine for cell growth extracted from the HAD model, $NN_{Gln,a}$ (cf. Table 1, row 6), is compared against μ_{gln} in eq 5 of the in silico model; (ii) the rate of glutamine consumption through cell-independent mechanisms extracted from the HAD model, $NN_{Gln,b}$ (cf. Table 1, row 6), is compared against $-r_{GLN}$ in eq 5; and (iii) the rate of NH_4 production from cell-independent pathways extracted from the HAD model, $NN_{NH_4,b}$ (cf. Table 1, row 6), is compared against r_{GLN} in eq 6. It is noted that the in silico model comprises a single term r_{GLN} in the equations of both GLN and NH_4. However, in the hybrid model, this is learned by two different NN outputs ($NN_{Gln,b}$ and $NN_{NH_4,b}$) for the two process variables GLN and NH_4, respectively. Correspondingly, two different comparisons are presented in Figure 5E,F.

It can be observed that, in all cases, the NN in the hybrid models could successfully learn the kinetic behavior from the in silico model by merely fitting the time-course data. Additionally, for most of the cases (except μ_d) the standard deviation in the extracted information is also low. Thus, hybrid models can be used to extract information about the specific rate that can highlight the differences in the metabolic behavior of the processes. Although the models in this example exhibit similar accuracy, the complex hybrid models (HXD and HAD) allow the extraction of useful and detailed information about the underlying process mechanisms, which can be used to improve process understanding by scientists. For instance, it is known that metabolites such as lactate and ammonia cause cell death, and it would be beneficial to keep them as low as possible. The profiles of r_{GLN} in Figure 5E,F indicate that in the considered system (process condition, type of media, etc.), there could be a possibility of the direct conversion of glutamine to ammonia. However, if the r_{GLN} values were to be close to zero, it would be an indication that in that particular system (dictated by process conditions, media type, etc.) this pathway is not active. Scientists can use the information extracted to select the process conditions (or media) leading to the desired behavior. Subsequently, depending on the specific underlying process we intend to investigate, one can select the most appropriate hybrid model. This is similar to the results reported in ref 56, in the case of downstream chromatographic processes, where different hybrid model variants have been developed that allow the understanding of mass transfer kinetics and (or) adsorption equilibria.

Of course, although potentially very useful, this approach should be considered with care. Due to the intrinsic structure of the hybrid models, all of the effects that are not described in the mechanistic part are accounted for in a lumped form in the data-driven portion of the model. Thus, these are defined as correlations, and the possibility of inferring causation must be handled with care and requires validation from independent process knowledge.

4. DISCUSSION

In this work, we could quantitatively demonstrate that, for cell culture processes, either extreme in the degree of hybridization, data-driven (0%) or mechanistic (100%) modeling, has limitations. The data-driven model performs poorly in interpolation and extrapolation when sufficient data is not provided. The increase in the degree of hybridization has been observed to be generally beneficial in reducing the cost of the model training and increasing its predictive accuracy, as long as the introduced process knowledge is not such that a bias is introduced in the model limiting general behavior prediction and extrapolation. The mechanistic model has the negative effect of excessively constraining the model, thus reducing its flexibility and the accuracy of its predictions. The most relevant

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

aspects are summarized in Table 2 to provide a systematic view of the conclusion of this work.

While transitioning from the data-driven to the mechanistic model, adding process knowledge improved the performance of the model progressively. Among the different pieces of knowledge, accounting for the feed strategy in the form of mass balances and the introduction of specific metabolic rates proved to be crucial. It has been found that there exists a trade-off between adding process knowledge and introducing more parameters in the model structure, thus requiring more experiments to train a robust model. To best exploit the features of hybrid models for process applications such as optimization, monitoring, and control, we need to identify the most convenient level of hybridization, which varies for different processes and applications. In the case investigated in this work, the HR model denotes the optimal compromise between adding process knowledge and increasing the model parameters, as well as performing best across various aspects, especially when only a limited number of data are available.

However, hybrid models with other degrees of hybridization could be helpful for various other applications. For instance, with the drift toward continuous manufacturing and corresponding switch from fed-batch to perfusion bioreactors, the ability to transfer models across different modes of operation could be attractive. For this purpose, hybrid models with mass-balance constraints (Hyb, HR, HXD, and HAD) have been shown to be preferable. On the other hand, hybrid models with a smaller degree of hybridization, such as Hyb-No MB, could be a better choice of a model when the available feeding strategy has some discrepancies.

Another possible consideration is that the different hybrid models can be used to extract macroscopic metabolic parameters such as the specific cell growth rate and the consumption or production rates of various metabolites at different levels of detail. Such information could provide a better understanding that can be used for the rational design and development of the process and is in line with the emphasis of the QbD initiative. In this regard, HXD and HAD types of models could be beneficial.

Additionally, this new concept can not only be used to extract process understanding but, in fact, also to test hypotheses. Especially for cell cultures, where the underlying phenomena (e.g., cell metabolism) are only partially understood, several hypotheses about the macro-kinetic models can be formulated. In particular, such hypotheses could be introduced sequentially, from the most likely to the least likely ones, to strategically test for the validity and identify breakpoints.

Although one could speculate about observing qualitatively similar results for many different processes, we should be careful and consider the above results valid for the system here considered. On the other hand, the procedure for determining the optimal degree of hybridization when using hybrid models can be regarded as general and is expected to provide a new degree of freedom for optimizing our modeling capabilities. Indeed, this has been confirmed in the case of the chromatographic processes used for the purification of therapeutic proteins, as illustrated in ref 56. Also, in this case, the degree of hybridization has been proven to provide an additional degree of freedom to optimize model performance.

5. CONCLUSIONS

In this work, we have demonstrated that hybrid models are better options than purely data-driven or mechanistic models for cell culture. However, hybrid models can be designed with different degrees of hybridization, i.e., with different levels of process knowledge embedded, and an appropriate choice among such models must be made based on data availability and the targeted application. In the case of cell culture applications, the underlying elementary processes are poorly understood. Especially the effect and interaction of the large number of process parameters (commonly observed in industrial practice). In such cases, hybridization can be used as a tool to test mechanistic hypotheses about the elementary processes and to enhance our level of process understanding on key process features. For instance, the cell-specific growth rate and how such intrinsic features are affected by process parameters.

In this work, only titer was considered in the modeling related to the product. However, there are several crucial critical quality attributes that are relevant. In the future, the effect of using different hybrid models in the CQA prediction must also be considered. In addition to the ones considered, there are other process factors such as media and clones that should be studied. Furthermore, only accuracy and process application-related metrics are evaluated in this work. However, there could be additional evaluation metrics that can be considered, for instance, ease of availability of knowledge, the expertise required to set up the model, and the computational time. While the former two are more qualitative discussions, the latter can be studied. However, with advanced packages, e.g., TorchDiffEq.[66] in python and DiffEqFlux[67] in Julia, reasonable computational speed can be achieved.

Finally, it is to be noted that the generality of this hybrid modeling approach can be appreciated in our previous work on this subject,[56] which focused on chromatographic protein purification processes but arrived at similar conclusions. Furthermore, the concept of hybrid modeling and degrees of hybridization can be applied to other domains (e.g., protein engineering or formulation development). For instance, in protein engineering problems,[68] a combination of molecular dynamics simulations and machine learning for engineering proteins could be explored following the similar concept of hybridization presented in this work.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.iecr.1c04507.

Data-driven model details and comparison (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Alessandro Butté − *DataHow AG, 8600 Dübendorf, Switzerland*; Email: a.butte@datahow.ch

Massimo Morbidelli − *DataHow AG, 8600 Dübendorf, Switzerland; Dipartimento di Chimica, Materiali e Ingegneria Chimica, Giulio Natta, Politecnico di Milano, 20131 Milano, Italy*; Email: massimo.morbidelli@polimi.it

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672

Authors

Harini Narayanan − *Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zurich, Switzerland;* orcid.org/0000-0003-4545-4885

Martin Luna − *Institute of Chemical and Bioengineering, Department of Chemistry and Applied Biosciences, ETH Zurich, 8093 Zurich, Switzerland*

Michael Sokolov − *DataHow AG, 8600 Dübendorf, Switzerland;* orcid.org/0000-0001-8396-4099

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.iecr.1c04507

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Prof. Paolo Arosio for providing valuable comments to improve the scientific presentation of this work.

■ REFERENCES

(1) Narayanan, H.; Luna, M. F.; von Stosch, M.; Cruz Bournazou, M. N.; Polotti, G.; Morbidelli, M.; Butté, A.; Sokolov, M. Bioprocessing in the Digital Age: The Role of Process Models. *Biotechnol. J.* **2020**, *15*, No. 1900172.

(2) Sokolov, M.; von Stosch, M.; Narayanan, H.; Feidl, F.; Butté, A. Hybrid Modeling — a Key Enabler towards Realizing Digital Twins in Biopharma? *Curr. Opin. Chem. Eng.* **2021**, *34*, No. 100715.

(3) Narayanan, H.; Sponchioni, M.; Morbidelli, M. Integration and Digitalization in the Manufacturing of Therapeutic Proteins. *Chem. Eng. Sci.* **2021**, *248*, No. 117159.

(4) Sokolov, M. Decision Making and Risk Management in Biopharmaceutical Engineering - Opportunities in the Age of Covid-19 and Digitalization. *Ind. Eng. Chem. Res.* **2020**, *59*, 17587−17592.

(5) von Stosch, M.; Davy, S.; Francois, K.; Galvanauskas, V.; Hamelink, J. M.; Luebbert, A.; Mayer, M.; Oliveira, R.; O'Kennedy, R.; Rice, P.; Glassey, J. Hybrid Modeling for Quality by Design and PAT-Benefits and Challenges of Applications in Biopharmaceutical Industry. *Biotechnol. J.* **2014**, *9*, 719−726.

(6) Glassey, J.; Gernaey, K. V.; Clemens, C.; Schulz, T. W.; Oliveira, R.; Striedner, G.; Mandenius, C. F. Process Analytical Technology (PAT) for Biopharmaceuticals. *Biotechnol. J.* **2011**, *6*, 369−377.

(7) Solle, D.; Hitzmann, B.; Herwig, C.; Pereira Remelhe, M.; Ulonska, S.; Wuerth, L.; Prata, A.; Steckenreiter, T. Between the Poles of Data-Driven and Mechanistic Modeling for Process Operation. *Chem. Ing. Tech.* **2017**, *89*, 542−561.

(8) Narayanan, H.; Luna, M.; Sokolov, M.; Arosio, P.; Butté, A.; Morbidelli, M. Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Capture Chromatographic Step. *Ind. Eng. Chem. Res.* **2021**, *60*, 10466−10478.

(9) Narayanan, H.; Sokolov, M.; Morbidelli, M.; Butté, A. A New Generation of Predictive Models: The Added Value of Hybrid Models for Manufacturing Processes of Therapeutic Proteins. *Biotechnol. Bioeng.* **2019**, *116*, 2540−2549.

(10) Pérez-Fernández, B. A.; Fernandez-de-Cossio-Diaz, J.; Boggiano, T.; León, K.; Mulet, R. In-Silico Media Optimization for Continuous Cultures Using Genome Scale Metabolic Networks: The Case of CHO-K1. *Biotechnol. Bioeng.* **2021**, *118*, 1884−1897.

(11) Opdam, S.; Richelle, A.; Kellman, B.; Li, S.; Zielinski, D. C.; Lewis, N. E. A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst.* **2017**, *4*, 318.e6−329.e6.

(12) Schinn, S. M.; Morrison, C.; Wei, W.; Zhang, L.; Lewis, N. E. Systematic Evaluation of Parameters for Genome-Scale Metabolic Models of Cultured Mammalian Cells. *Metab. Eng.* **2021**, *66*, 21−30.

(13) Carvalho, M.; Nikdel, A.; Riesberg, J.; Lyons, D.; Budman, H. Identification of a Dynamic Metabolic Flux Model for a Mammalian Cell Culture. *IFAC-PapersOnLine* **2019**, *52*, 88−93.

(14) Galleguillos, S. N.; Ruckerbauer, D.; Gerstl, M. P.; Borth, N.; Hanscho, M.; Zanghellini, J. What Can Mathematical Modelling Say about CHO Metabolism and Protein Glycosylation? *Comput. Struct. Biotechnol. J.* **2017**, *15*, 212−221.

(15) Richelle, A.; David, B.; Demaegd, D.; Dewerchin, M.; Kinet, R.; Morreale, A.; Portela, R.; Zune, Q.; von Stosch, M. Towards a Widespread Adoption of Metabolic Modeling Tools in Biopharmaceutical Industry: A Process Systems Biology Engineering Perspective. *npj Syst. Biol. Appl.* **2020**, *6*, No. 6.

(16) Martínez, J. A.; Bulté, D. B.; Contreras, M. A.; Palomares, L. A.; Ramírez, O. T. Dynamic Modeling of CHO Cell Metabolism Using the Hybrid Cybernetic Approach With a Novel Elementary Mode Analysis Strategy. *Front. Bioeng. Biotechnol.* **2020**, *8*, No. 279.

(17) Craven, S.; Shirsat, N.; Whelan, J.; Glennon, B. Process Model Comparison and Transferability across Bioreactor Scales and Modes of Operation for a Mammalian Cell Bioprocess. *Biotechnol. Prog.* **2013**, *29*, 186−196.

(18) Xing, Z.; Bishop, N.; Leister, K.; Li, Z. J. Modeling Kinetics of a Large-Scale Fed-Batch CHO Cell Culture by Markov Chain Monte Carlo Method. *Biotechnol. Prog.* **2010**, *26*, 208−219.

(19) Glassey, J.; von Stosch, M., Eds. *Hybrid Modeling in Process Industries*, 1st ed.; CRC Press, 2018.

(20) Thompson, M. L.; Kramer, M. A. Modeling Chemical Processes Using Prior Knowledge and Neural Networks. *AIChE J.* **1994**, *40*, 1328−1340.

(21) Schubert, J.; Simutis, R.; Dors, M.; Havlik, I.; Lubbert, A. Hybrid Modeling of Yeast Production Processes - Combination of a-Priori Knowledge on Different Levels of Sophistication. *Chem. Eng. Technol.* **1994**, *17*, 10−20.

(22) Psichogios, D. C.; Ungar, L. A Hybrid Neural Network-First Principles Approach to Process Modeling. *AIChE J.* **1992**, *38*, 337−346.

(23) Can, H. J. L.; Van; Braake, H. A. B.; Hellinga, C.; Luyben, K. C. A. M.; Heijnen, J. J. An Efficient Model Development Strategy for Bioprocesses Based on Neural Networks in Macroscopic Balances. *Biotechnol. Bioeng.* **1997**, *54*, 550−566.

(24) Xiong, Q.; Jutan, A. Grey-Box Modelling and Control of Chemical Processes. *Chem. Eng. Sci.* **2002**, *57*, 1027−1039.

(25) Mogk, G.; Mrziglod, T.; Schuppert, A. Application of Hybrid Models in Chemical Industry. *Computer Aided Chemical Engineering*; Elsevier, 2002; Vol. *10*.

(26) Bellos, G. D.; Kallinikos, L. E.; Gounaris, C. E.; Papayannakos, N. G. Modelling of the Performance of Industrial HDS Reactors Using a Hybrid Neural Network Approach. *Chem. Eng. Process.* **2005**, *44*, 505−515.

(27) Hu, G.; Mao, Z.; He, D.; Yang, F. Hybrid Modeling for the Prediction of Leaching Rate in Leaching Process Based on Negative Correlation Learning Bagging Ensemble Algorithm. *Comput. Chem. Eng.* **2011**, *35*, 2611−2617.

(28) Yang, A.; Martin, E.; Morris, J. Identification of Semi-Parametric Hybrid Process Models. *Comput. Chem. Eng.* **2011**, *35*, 63−70.

(29) Bhutani, N.; Rangaiah, G. P.; Ray, A. K. First-Principles, Data-Based, and Hybrid Modeling and Optimization of an Industrial Hydrocracking Unit. *Ind. Eng. Chem. Res.* **2006**, *45*, 7807−7816.

(30) Georgieva, P.; Meireles, M. J.; Feyo de Azevedo, S. Knowledge-Based Hybrid Modelling of a Batch Crystallisation When Accounting for Nucleation, Growth and Agglomeration Phenomena. *Chem. Eng. Sci.* **2003**, *58*, 3699−3713.

(31) Ghosh, D.; Hermonat, E.; Mhaskar, P.; Snowling, S.; Goel, R. Hybrid Modeling Approach Integrating First-Principles Models with Subspace Identification. *Ind. Eng. Chem. Res.* **2019**, *58*, 13533−13543.

(32) Tian, Y.; Zhang, J.; Morris, J. Modeling and Optimal Control of a Batch Polymerization Reactor Using a Hybrid Stacked Recurrent Neural Network Model. *Ind. Eng. Chem. Res.* **2001**, *40*, 4525−4535.

(33) Winkelmann, S.; Schütte, C. Hybrid Models for Chemical Reaction Networks: Multiscale Theory and Application to Gene Regulatory Systems. *J. Chem. Phys.* **2017**, *147*, No. 114115.
(34) Zander, H.-J.; Dittmeyer, R.; Wagenhuber, J. Dynamic Modeling of Chemical Reaction Systems with Neural Networks and Hybrid Models. *Chem. Eng. Technol.* **1999**, *21*, 571−574.
(35) Nagrath, D.; Messac, A.; Bequette, B. W.; Cramer, S. M. A Hybrid Model Framework for the Optimization of Preparative Chromatographic Processes. *Biotechnol. Prog.* **2008**, *20*, 162−178.
(36) Zhang, J.; Mao, Z.-z.; Jia, R.-d.; He, D.-k. Real Time Optimization Based on a Serial Hybrid Model for Gold Cyanidation Leaching Process. *Miner. Eng.* **2015**, *70*, 250−263.
(37) Anderson, J. S.; McAvoy, T. J.; Hao, O. J. Use of Hybrid Models in Wastewater Systems. *Ind. Eng. Chem. Res.* **2000**, *39*, 1694−1704.
(38) Zendehboudi, S.; Rezaei, N.; Lohi, A. Applications of Hybrid Models in Chemical, Petroleum, and Energy Systems: A Systematic Review. *Appl. Energy* **2018**, *228*, 2539−2566.
(39) Chen, L.; Bernard, O.; Bastin, G.; Angelov, P. Hybrid Modelling of Biotechnological Processes Using Neural Networks. *Control Eng. Pract.* **2000**, *8*, 821−827.
(40) Feyo de Azevedo, S.; Dahm, B.; Oliveira, F. R. Hybrid Modelling of Biochemical Processes: A Comparison with the Conventional Approach. *Comput. Chem. Eng.* **1997**, *21*, S751−S756.
(41) Zorzetto, L. F. M.; Maciel Filho, R.; Wolf-Maciel, M. R. Process Modelling Development through Artificial Neural Networks and Hybrid Models. *Comput. Chem. Eng.* **2000**, *24*, 1355−1360.
(42) Narayanan, H.; Cruz Bournazou, M. N.; Guillén Gosálbez, G.; Butté, A. Functional-Hybrid Modeling through Automated Adaptive Symbolic Regression for Interpretable Mathematical Expressions. *Chem. Eng. J.* **2021**, No. 133032.
(43) Hutter, C.; von Stosch, M.; Cruz Bournazou, M. N.; Butté, A. Knowledge Transfer across Cell Lines Using Hybrid Gaussian Process Models with Entity Embedding Vectors. *Biotechnol. Bioeng.* **2021**, *118*, 4389−4401.
(44) von Stosch, M.; Hamelink, J.-M.; Oliveira, R. Hybrid Modeling as a QbD/PAT Tool in Process Development: An Industrial E. coli Case Study. *Bioprocess Biosyst. Eng.* **2016**, *39*, 773−784.
(45) Teixeira, A.; Cunha, A. E.; Clemente, J. J.; Moreira, J. L.; Cruz, H. J.; Alves, P. M.; Carrondo, M. J. T.; Oliveira, R. Modelling and Optimization of a Recombinant BHK-21 Cultivation Process Using Hybrid Grey-Box Systems. *J. Biotechnol.* **2005**, *118*, 290−303.
(46) Von Stosch, M.; Oliveria, R.; Peres, J.; De Azevedo, S. F. Hybrid Modeling Framework for Process Analytical Technology: Application to Bordetella Pertussis Cultures. *Biotechnol. Prog.* **2012**, *28*, 284−291.
(47) Teixeira, A. P.; Alves, C.; Alves, P. M.; Carrondo, M. J.; Oliveira, R. Hybrid Elementary Flux Analysis/Nonparametric Modeling: Application for Bioprocess Control. *BMC Bioinf.* **2007**, *8*, No. 30.
(48) Teixeira, A. P.; Clemente, J. J.; Cunha, A. E.; Carrondo, M. J. T.; Oliveira, R. Bioprocess Iterative Batch-to-Batch Optimization Based on Hybrid Parametric/Nonparametric Models. *Biotechnol. Prog.* **2006**, *22*, 247−258.
(49) Montague, G. A.; Glassey, J.; Ignova, M.; Paul, G. C.; Kent, C. A.; Thomas, C. R.; Ward, A. C. Hybrid Modelling for On-Line Penicillin Fermentation Optimisation. *IFAC Proc. Vol.* **2002**, *35*, 395.
(50) Schubert, J.; Simutis, R.; Dors, M.; Havlik, I.; Lübbert, A. Bioprocess Optimization and Control: Application of Hybrid Modelling. *J. Biotechnol.* **1994**, *35*, 51−68.
(51) Oliveira, R.; Simutis, R.; Feyo De Azevedo, S.; Lübbert, A. HYBNET, An Advanced Tool for Process Optimization and Control. *IFAC Proc. Vol.* **1998**, *31*, 289−294.
(52) Zorzetto, L. F. M.; Wilson, J. A. Monitoring Bioprocesses Using Hybrid Models and an Extended Kalman Filter. *Comput. Chem. Eng.* **1996**, *20*, S689−S694.
(53) Galvanauskas, V.; Simutis, R.; Lbbert, A. Hybrid Process Models for Process Optimisation, Monitoring and Control. *Bioprocess Biosyst. Eng.* **2004**, *26*, 393−400.
(54) Narayanan, H.; Behle, L.; Luna, M. F.; Sokolov, M.; Guillén-Gosálbez, G.; Morbidelli, M.; Butté, A. Hybrid-EKF: Hybrid Model Coupled with Extended Kalman Filter for Real-Time Monitoring and Control of Mammalian Cell Culture. *Biotechnol. Bioeng.* **2020**, *117*, 2703−2714.
(55) von Stosch, M.; Davy, S.; Francois, K.; Galvanauskas, V.; Hamelink, J. M.; Luebbert, A.; Mayer, M.; Oliveira, R.; O'Kennedy, R.; Rice, P.; Glassey, J. Hybrid Modeling for Quality by Design and PAT-Benefits and Challenges of Applications in Biopharmaceutical Industry. *Biotechnol. J.* **2014**, *9*, 719−726.
(56) Narayanan, H.; Luna, M.; Sokolov, M.; Arosio, P.; Butté, A.; Morbidelli, M. Hybrid Models Based on Machine Learning and an Increasing Degree of Process Knowledge: Application to Capture Chromatographic Step. *Ind. Eng. Chem. Res.* **2021**, *60*, 10466.
(57) Sokolov, M.; Ritscher, J.; MacKinnon, N.; Souquet, J.; Broly, H.; Morbidelli, M.; Butté, A. Enhanced Process Understanding and Multivariate Prediction of the Relationship between Cell Culture Process and Monoclonal Antibody Quality. *Biotechnol. Prog.* **2017**, *33*, 1368−1380.
(58) Rouiller, Y.; Solacroup, T.; Deparis, V.; Barbafieri, M.; Gleixner, R.; Broly, H.; Eon-Duval, A. Application of Quality by Design to the Characterization of the Cell Culture Process of an Fc-Fusion Protein. *Eur. J. Pharm. Biopharm.* **2012**, *81*, 426−437.
(59) Narayanan, H.; Sokolov, M.; Butté, A.; Morbidelli, M. Decision Tree-PLS (DT-PLS) Algorithm for the Development of Process: Specific Local Prediction Models. *Biotechnol. Prog.* **2019**, *35*, No. e2818.
(60) Sokolov, M.; Soos, M.; Neunstoecklin, B.; Morbidelli, M.; Butté, A.; Leardi, R.; Solacroup, T.; Stettler, M.; Broly, H. Fingerprint Detection and Process Prediction by Multivariate Analysis of Fed-Batch Monoclonal Antibody Cell Culture Data. *Biotechnol. Prog.* **2015**, *31*, 1633−1644.
(61) Narayanan, H.; Sokolov, M.; Butté, A.; Morbidelli, M. Decision Tree-PLS (DT-PLS) Algorithm for the Development of Process: Specific Local Prediction Models. *Biotechnol. Prog.* **2019**, *35*, 1−11.
(62) Jang, J. D.; Barford, J. P. An Unstructured Kinetic Model of Macromolecular Metabolism in Batch and Fed-Batch Cultures of Hybridoma Cells Producing Monoclonal Antibody. *Biochem. Eng. J.* **2000**, *4*, 153−168.
(63) Tatiraju, S.; Soroush, M.; Mutharasan, R. In *Multi-Rate Nonlinear State and Parameter Estimation in a Bioreactor*, Proceedings of the 1998 American Control Conference. ACC (IEEE Cat. No.98CH36207), 1998; pp 2324−2328.
(64) Von Stosch, M.; Oliveira, R.; Peres, J.; Feyo De Azevedo, S. A Novel Identification Method for Hybrid (N)PLS Dynamical Systems with Application to Bioprocesses. *Expert Syst. Appl.* **2011**, *38*, 10862−10874.
(65) Duarte, B. P. M.; Saraiva, P. M. Hybrid Models Combining Mechanistic Models with Adaptive Regression Splines and Local Stepwise Regression. *Ind. Eng. Chem. Res.* **2003**, *42*, 99−107.
(66) Chen, R.; Rubanova, Y.; Bettencourt, J.; Duvenaud, D. Neural Ordinary Differential Equations Background: ODE Solvers. *Neural Inf. Process. Syst.* **2019**, 6571−6583.
(67) Rackauckas, C.; Ma, Y.; Martensen, J.; Warner, C.; Zubov, K.; Supekar, R.; Skinner, D.; Ramadhan, A.; Edelman, A. Universal Differential Equations for Scientific Machine Learning, 2020; pp 1−55.
(68) Narayanan, H.; Dingfelder, F.; Butté, A.; Lorenzen, N.; Sokolov, M.; Arosio, P. Machine Learning for Biologics: Opportunities for Protein Engineering, Developability, and Formulation. *Trends Pharmacol. Sci.* **2021**, *42*, 151−165.

https://doi.org/10.1021/acs.iecr.1c04507
Ind. Eng. Chem. Res. 2022, 61, 8658−8672