






RESEARCH ARTICLE

Cell Culture and Tissue Engineering

BIOTECHNOLOGY
PROGRESS

Efficient optimization of time-varying inputs in a fed-batch cell culture process using design of dynamic experiments

Yu Luo¹  | Duane A. Stanton² | Rachel C. Sharp¹  | Alexis J. Parrillo¹  |
Kelsey T. Morgan¹  | Diana B. Ritz¹ | Sameer Talwar¹ ¹GSK, Biopharm Drug Substance Development, King of Prussia, Pennsylvania, USA²GSK, CMC Statistics, King of Prussia, Pennsylvania, USA**Correspondence**

Yu Luo, GSK, Biopharm Drug Substance Development, King of Prussia, Pennsylvania, USA.

Email: yu.8.luo@gsk.com**Abstract**

In cell culture process development, we rely largely on an iterative, one-factor-at-a-time procedure based on experiments that explore a limited process space. Design of experiments (DoE) addresses this issue by allowing us to analyze the effects of process inputs on process responses systematically and efficiently. However, DoE cannot be applied directly to study time-varying process inputs unless an impractically large number of bioreactors is used. Here, we adopt the methodology of design of dynamic experiments (DoDE) and incorporate dynamic feeding profiles efficiently in late-stage process development of the manufacture of therapeutic monoclonal antibodies. We found that, for the specific cell line used in this article, (1) not only can we estimate the effect of nutrient feed amount on various product attributes, but we can also estimate the effect, develop a statistical model, and use the model to optimize the slope of time-trended feed rates; (2) in addition to the slope, higher-order dynamic characteristics of time-trended feed rates can be incorporated in the design but do not have any significant effect on the responses we measured. Based on the DoDE data, we developed a statistical model and used the model to optimize several process conditions. Our effort resulted in a tangible improvement in productivity—compared with the baseline process without dynamic feeding, this optimized process in a 200-L batch achieved a 27% increase in titer and > 92% viability. We anticipate our application of DoDE to be a starting point for more efficient workflows to optimize dynamic process conditions in process development.

1 | INTRODUCTION

It is one of the major challenges faced by the biopharmaceutical industry to achieve desired productivity and product quality consistently.¹ Therapeutic monoclonal antibodies (mAbs) are typically manufactured using Chinese hamster ovary (CHO) cell cultures in a controlled fed-batch bioreactor environment. Process conditions and the feeding strategy for nutrient supplementation critically affect the growth and metabolic behaviors of cells, which in turn determine the final product titer and quality. While we still rely largely on an iterative, one-factor-at-a-time (OFAT) procedure based on experiments

that explore a limited process space, model-based process optimization is widely adopted in process development in other industries² and, with an increasing prevalence, in developing bioprocesses.^{3–7} Response surface methodology (RSM) models, for example, are suitable for model-based cell culture process optimization as they can be used to predict the final productivity or product quality responses based on process inputs such as temperature, pH, and other process conditions.^{6,8} Creating an RSM model requires data from properly designed experiments such that the effects of process inputs on productivity and product quality responses can be estimated systematically.

Design of experiments (DoE)⁹ is suitable for designing experiments systematically with *static* process conditions that vary only across batches, not in time. Typically, a design consists of several factors, and their levels are varied simultaneously. Each factor represents a process condition (or configuration) of interest and varies between low (−1) and high (+1) levels. As orthogonality and statistical criteria such as D-optimality are often satisfied in a DoE, the resulting design can be used to explore an experimental space more efficiently than an iterative, OFAT design, given the same constraint of resources, for example, a limited number of batches.

The conventional DoE approach, however, is not capable of describing time-varying, *dynamic* process conditions. A naïve modification would be creating multiple variable–time composite factors (e.g., pH–Day 0, pH–Day 1, pH–Day 2, etc.); however, not only does this design violate the orthogonality requirement, but the design can also become impossibly large and exceed the bioreactor limit of a typical experiment run in a well-equipped process development laboratory. Alternatively, if we treat a time-dynamic profile as a categorical factor and consider, for example, only linear trajectories, we can create a design of $3^2 = 9$ potential trajectories that start low/mid/high and end low/mid/high. Adding three center points will result in a total of 27 runs (each center point is replicated for any given categorical level to maintain design balance). If we add more factors, then the already large number of runs will quickly become even larger and impractical to implement.

In Section 2, we describe the design of *dynamic* experiments (DoDE) approach¹⁰ to designing experiments with time-varying process inputs *efficiently* (see References 6,10–12 for applications of DoDE to different processes). This approach allows us to study the effects of both static process inputs (e.g., constant temperature and pH setpoints) and dynamic process inputs (e.g., feed rates) in an orthogonal cell culture process development experiment, while keeping the number of conditions manageable. In this work, we focus on analyzing the time-varying rates of cell-specific feeding in fed-batch processes.

In a typical fed-batch process, supplemental nutrients are added to the bioreactor in either boli or a continuous flow because nutrients from the basal medium alone may be depleted. Both the composition and the added amount of the feed medium can affect productivity and product quality.^{13–15} In this work, we focus only on the optimization of the continuous feed rates. There are several ways of determining the amount to feed each day. Under cell-specific feeding, the amount of a nutrient feed medium added into a bioreactor is proportional to the current viable cell concentration (VCC) based on a pre-determined ratio–cell-specific feed rate (CSFR).¹⁶ This approach allows one to account for the increasing or decreasing demand for nutrients as cells grow and die. An optimal CSFR setpoint is crucial to the success of a process that uses cell-specific feeding. Normally, the CSFR setpoint of a process remains constant based on the assumption that each cell consumes nutrients at a constant rate throughout a fed-batch process. However, the metabolism of cells can shift substantially over the course of a fed-batch process, particularly as cell viability declines. Therefore, the optimal CSFR setpoint should account for

such shifts. Optimizing a time-varying CSFR setpoint trend (instead of a constant CSFR setpoint) is the focus of this article.

The remainder of this article is organized as follows. In Section 2, we review the DoDE methodology and describe the experimental details. In Section 3, we discuss the designs, data, and models of two small-scale experiments and the optimization of process conditions to improve product titer. Specifically, we estimated the effects, developed a statistical model, and used the model to optimize both the average amount and the slope of time-trended CSFRs; we verified the optimized process in a large-scale experiment and saw improved titer and cell viability; in addition to the slope, the curvature of time-trended CSFRs was also incorporated in a DoDE and found not have any significant effect on the responses we measured. In Section 4, we summarize our findings and discuss potential challenges and future direction.

2 | METHODS

The DoDE methodology introduced in Reference 10 is used in this article to design experiments with time-varying CSFR setpoints. Each CSFR trend can be described by one or more DoDE factors, depending on their respective levels. DoDE factors follow the same format as DoE factors where each factor's (coded) level falls between −1 (low level) and +1 (high level).

Orthogonal basis functions of time represent different and independent time-varying components of a CSFR trend. The absolute value of each DoDE factor level represents the “weight” of its corresponding basis function. A defined CSFR trend is the weighted sum of these basis functions. Specifically, the following three shifted Legendre polynomials form the bases of a quadratic CSFR trend:

$$P_0(\tau) = 1; P_1(\tau) = -1 + 2\tau; P_2(\tau) = 1 - 6\tau + 6\tau^2,$$

where P_i denotes an i th-order Legendre polynomial of time (bounded between −1 and +1); τ denotes the normalized time (bounded between 0 and 1). Higher-order Legendre polynomials are also available to describe more complex trends. The Legendre polynomials are orthogonal to each other according to their zero inner products.

$$\int_0^1 P_i(\tau)P_j(\tau) d\tau = 0 \quad (i \neq j).$$

In other words, when a time-varying trend consists of multiple Legendre polynomials, one can adjust the component of the trend that corresponds to a specific Legendre polynomial *independently* and analyze its effect on the responses in isolation.

A normalized time-varying CSFR trend is represented as follows:

$$z(\tau) = x_1P_0(\tau) + x_2P_1(\tau) + x_3P_2(\tau),$$

where z is the normalized CSFR trend; x_1 , x_2 , and x_3 are the DoDE factor levels (or weights). Note that in Section 3, x_1 , x_2 , and x_3 are

replaced by the corresponding factor names (e.g., CSFR₀, CSFR₁, and CSFR₂). z should also be bounded by -1 (the lowest allowed CSFR) and $+1$ (the highest allowed CSFR); however, with $-1 \leq P_i \leq 1$ and $-1 \leq x_i \leq 1$, it is not guaranteed that z also falls within ± 1 . For example, if $x_1 = x_2 = x_3 = 1$, $z(1) = 1 + 1 + 1 = 3 > 1$. A remedy without changing the relative level magnitudes of the design is to scale x_i by the number of factors such that $|z| \leq 1$. In the same example, the scaled factors are $x_1 = x_2 = x_3 = \frac{1}{3}$ when the three factor levels are high. Typically, the execution of a theoretical CSFR trend, z , in an experiment requires that z is discretized into finite steps. An operator or automation program then updates the feed rate periodically.

Designing an experiment using DoDE is mostly identical to designing an experiment using DoE as follows:

1. Determine the static DoE factors and the DoDE factors of a study.
 - a. Determine the number of feasible runs (same as DoE).
 - b. Determine the lower and upper limits of the static conditions described by the DoE factors and translate them into coded factor levels (same as DoE).
 - c. Determine the lower and upper limits of the time-varying conditions (e.g., CSFR) described by the DoDE factors and translate them into coded factor levels.
 - d. Determine the capacity to adjust the level of any time-varying conditions: e.g., linear trajectories or quadratic trajectories; coarse discretization (updating once per batch) or fine discretization (updating daily).
 - e. Select a suitable design (e.g., full factorial, fractional factorial, central composite, or optimal designs) based on the number of bioreactors that can feasibly be run per study (same as DoE).
2. Execute the study and collect the data (same as DoE).
3. Perform model/variable selection to identify the main and interaction effects for analysis of variance (ANOVA) and creating a linear model for each response variable (same as DoE).
 - a. Optimize both static and time-varying conditions subject to constraints based on the models developed from the last step (same as DoE except for the additional constraint of $\sum_i |x_i| \leq 1$ in DoDE).
 - b. Translate the optimized, coded DoE factor levels into actual conditions (same as DoE).
 - c. Translate the optimized, coded DoDE factor levels into discrete time steps.

The effect of dynamic CSFR strategies on product titer is estimated and analyzed in two sequential DoDE studies in this article. The objective of both studies is to maximize product titer as part of a fed-batch cell culture process development campaign. The first study serves as an exploratory exercise to screen CSFR-related factors and other static operating conditions. The second study builds upon the insights from the first study to validate and fine-tune the optimal CSFR strategy.

Analyzing the data from a DoDE study is identical to analyzing the data from a conventional DoE study. Common techniques such as ANOVA also apply to DoDE. In the bioreactor studies in the next section, we employed RSM to optimize CSFR and other process conditions.

$$y = \bar{y} + c_1x_1 + c_2x_2 + c_3x_3 + c_{11}x_1^2 + c_{22}x_2^2 + c_{33}x_3^2 + c_{12}x_1x_2 + c_{23}x_2x_3 + c_{13}x_1x_3,$$

where y represents the response variable; \bar{y} is the fixed effect; c is the coefficient of a main or interaction effect estimated based on data. Only quadratic main effects and two-way interactions are included in the equation above as higher-order effects are relatively rare (i.e., the hierarchical ordering principle).

The factors are selected in an RSM model based on standard factor selection criteria—statistical significance of the effects, adjusted R^2 , prediction R^2 , and preservation of the factor hierarchy (e.g., if interaction x_1x_2 is selected in a model, their corresponding main effects, x_1 and x_2 , then must also be selected even if they are not statistically significant). All RSM models in this article were created in Design-Expert 10.0.3 (Stat-Ease, Minneapolis, MN). We use adjusted R^2 (R_{adj}^2) as the selection criterion in Design-Expert to select the effect terms in all RSM models, while preserving the hierarchy of factors. Further refinement of effect term selection may be performed if needed.

The optimization of process conditions follows a standard constrained optimization procedure where one response variable (e.g., the final titer) is maximized (or minimized) subject to several constraints. All model-based process condition optimizations were solved using the Solver add-on in Microsoft Excel (Microsoft, Redmond, WA).

All DoDE experiments were carried out in an Ambr250 bioreactor system (Sartorius Stedim, Göttingen, Germany), and all large-scale experiments were carried out in a 200-L single-use bioreactor system (Sartorius Stedim, Göttingen, Germany). A CHO K1 glutamine synthetase (GS)-knockout cell line expressing an IgG1 monoclonal antibody was used. Proprietary and chemically defined seed expansion medium, production basal medium, and production feed medium were used. The initial working volume was 200 mL (Ambr250 DoDE batches) or 160–170 L (large-scale batches). The target seeding density was 18 million cells/mL. The pH was set between 6.9 ± 0.1 and 7.1 ± 0.1 . The temperature was kept between 32°C and 34°C . The agitation rate was determined for different scales based on a constant power per unit volume value (P/V), and the gas flow was determined similarly based on a constant volume of air sparged per unit volume per minute (VVM). The reactors were agitated at 435 rpm (DoDE batches) or 83 rpm (large-scale batches). The dissolved oxygen (DO) was controlled at 50%. The feed rate of each reactor was determined based on the specified CSFR and the VCC measured every day in the equation below.

$$F_{\text{feed}} = \text{CSFR} \times x_v \times V,$$

where F_{feed} (mL/h) is the volumetric flowrate of the nutrient feed, CSFR (pL/cell/min) is the specified CSFR, x_v (cells/mL) is the measured VCC, and V (L) is the working volume of the bioreactor. A 400 g/L glucose solution was used to provide supplemental glucose as needed. All experiments were run for 12 days.

A sample of the culture was taken daily to measure the antibody titer, cell densities, and metabolite concentrations in each bioreactor. The antibody titer was measured using an Agilent 1290 Infinity II LC system (Agilent, Santa Clara, CA). The cell densities and cell culture

viability were measured using a Vi-Cell XR cell counter (Beckman, Indianapolis, IN). The metabolite concentrations were measured using a BioProfile FLEX2 analyzer (Nova Biomedical, Waltham, MA) or a Cedex Bio HT analyzer (Roche, Mannheim, Germany).

3 | RESULTS AND DISCUSSION

Two Ambr250 studies—Study A and Study B—were designed using the DoDE methodology and executed consecutively to generate the data for the RSM model development and model-based optimization.

In the initial Ambr250 study (Study A), a DoDE was created to estimate the effects of dynamic CSFR, temperature, and pH on the end-of-run (EoR) mAb titer and cell viability as those process inputs are known to affect cell growth and protein production.^{17–20} Table 1 shows both the design and the measured responses. It is a 2^4 full-factorial design with four center points (the design is orthogonal). Each CSFR trend was dynamic in the experiment and described by two DoDE factors—CSFR₀ for the average feed rate and CSFR₁ for the speed of feed rate change or the slope (the trend was bounded between CSFR_{min} = 0.0002 pL/cell/min and CSFR_{max} = 0.0004 pL/cell/min based on the historical, static CSFRs used in the previous, iterative process development studies) in the equations below. DoDE was meant for further “fine-tuning” of the process, and the CSFR bounds were determined such that the conditions tested in this study would not result in a complete batch failure due to over- or under-feeding.

$$z(\tau) = \text{CSFR}_0 P_0(\tau) + \text{CSFR}_1 P_1(\tau) = \text{CSFR}_0 + \text{CSFR}_1(-1 + 2\tau),$$

$$\tau = \frac{t}{12},$$

where $-1 \leq z \leq 1$ and $0 \leq \tau \leq 1$ represent the normalized CSFR and time; $-1 \leq \text{CSFR}_0 \leq 1$ and $-1 \leq \text{CSFR}_1 \leq 1$ are the two (coded) DoDE factors. The CSFR at any time point, $z(\tau)$, is a linear combination of $P_0(\tau)$, which is a constant, and $P_1(\tau)$, which is an either rising or declining slope. We can obtain the simplest set of DoDE conditions by using at least two CSFR factors—CSFR₀ (average) and CSFR₁ (slope).

Specifically, the normalized CSFR is defined in the scaling equation below. To obtain a CSFR curve based on the DoDE factor levels, one calculates the normalized CSFR curve, $z(\tau)$, as a function of τ and recovers CSFR to its original scale based on the normalized values.

$$z = \frac{\text{CSFR} - \frac{1}{2}(\text{CSFR}_{\min} + \text{CSFR}_{\max})}{\text{CSFR}_{\max} - \frac{1}{2}(\text{CSFR}_{\min} + \text{CSFR}_{\max})}.$$

CSFR is normalized between CSFR_{min} and CSFR_{max}, and t is normalized by the duration of each batch (i.e., 12 days). To ensure z is bounded by ± 1 , CSFR₀ and CSFR₁ were scaled such that the space of allowable CSFR conditions in this design was bounded by $-0.5 \leq \text{CSFR}_0 \leq 0.5$ and $-0.5 \leq \text{CSFR}_1 \leq 0.5$. Alternatively, one can

TABLE 1 The design and responses of Study A.

Run order	CSFR ₀	CSFR ₁	T	pH	Titer (g/L)	Via (%)
1	0.5	0.5	1	-1	6.71	69.93
2	-0.5	-0.5	1	1	6.58	82.33
3	-0.5	0.5	-1	-1	5.88	75.16
4	0.5	0.5	-1	-1	5.34	74.71
5	0	0	0	0	7.23	81.25
6	0.5	-0.5	-1	-1	5.12	70.78
7	0.5	0.5	-1	1	6.53	79.35
8	0.5	-0.5	-1	1	6.48	76.99
9	-0.5	-0.5	-1	1	6.34	84.84
10	-0.5	0.5	-1	1	5.76	87.44
11	0	0	0	0	6.57	79.40
12	0.5	-0.5	1	-1	6.37	59.02
13	-0.5	0.5	1	1	6.52	81.98
14	0	0	0	0	7.38	80.27
15	-0.5	-0.5	1	-1	6.20	58.20
16	0	0	0	0	6.48	76.17
17	-0.5	-0.5	-1	-1	5.68	74.37
18	0.5	0.5	1	1	8.34	84.21
19	0.5	-0.5	1	1	8.35	77.16
20	-0.5	0.5	1	-1	6.03	76.18

assign CSFR₀ and CSFR₁ the coded levels of $(\pm 1, 0)$ and $(0, \pm 1)$ to cover a larger space while still satisfying the constraint that z is bounded by ± 1 . Due to the limited number of runs (20 bioreactors), we did not include such axial points in Study A.

Below is a step-by-step breakdown of translating the coded CSFR factor levels into the actual CSFR levels using Run 6 in Table 1 as an example:

1. Calculate the normalized time series.

$$\tau = \frac{t}{12} = \frac{[0, 1, \dots, 11, 12]}{12} = [0, 0.0833, \dots, 0.917, 1].$$

2. Calculate the normalized CSFR.

$$z = \text{CSFR}_0 + \text{CSFR}_1(-1 + 2\tau) = 0.5 + (-0.5) \times (-1 + 2\tau) = [1, 0.917, \dots, 0.0833, 0].$$

3. Substitute z , CSFR_{min}, and CSFR_{max} in the scaling equation to calculate the actual CSFR levels.

$$\begin{aligned} \text{CSFR} &= z \times \left[\text{CSFR}_{\max} - \frac{1}{2}(\text{CSFR}_{\min} + \text{CSFR}_{\max}) \right] \\ &\quad + \frac{1}{2}(\text{CSFR}_{\min} + \text{CSFR}_{\max}) \\ &= z \times \left[0.0004 - \frac{1}{2}(0.0002 + 0.0004) \right] + \frac{1}{2}(0.0002 + 0.0004) \\ &= [0.0004, 0.000392, \dots, 0.000308, 0.0003]. \end{aligned}$$

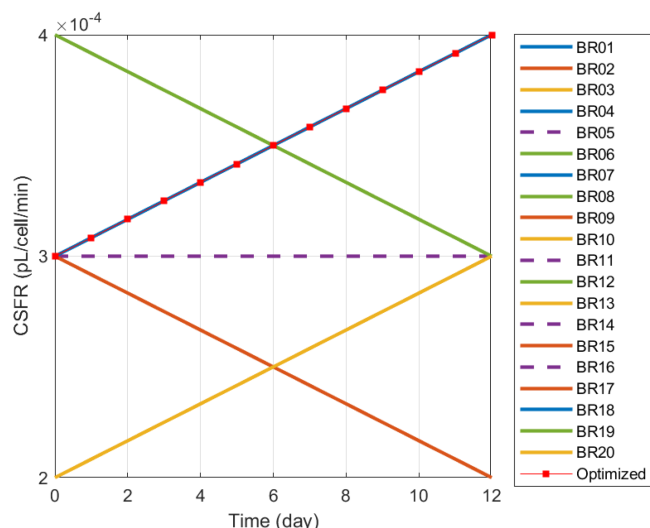


FIGURE 1 Theoretical CSFR trends in Study A. “BR” stands for “bioreactor” and is numbered according to the Run numbers in Table 1.

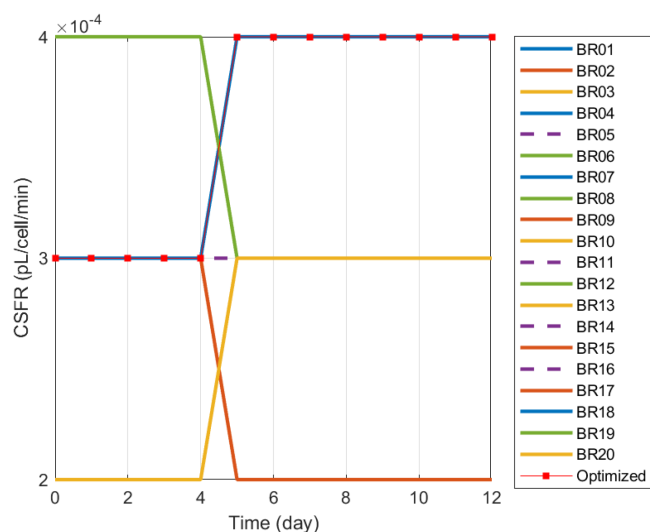


FIGURE 2 CSFR trends executed in Study A.

The temperature setpoint was static and described by a conventional DoE factor— T (bounded between low temperature at $T_{low} = 32^{\circ}\text{C}$ and high temperature at $T_{high} = 34^{\circ}\text{C}$); the pH setpoint was also static and described by another conventional DoE factor—pH (bounded between low pH at $\text{pH}_{low} = 6.9$ and high pH at $\text{pH}_{high} = 7.1$). The temperature and pH bounds were determined using the historical data from the previous development batches. All the factor levels in Table 1 are dimensionless, coded levels. All the CSFR trends are visualized in Figures 1 and 2. The change of CSFR is programmed in an Ambr250 system as a discrete event. In Study A, it was executed such that the CSFR changed once during each batch on Day 5—a coarse discretization of the smooth, theoretical trends in Figure 1, but the discretization does not affect the interpretation of results as the optimized CSFR trend was executed using the same

TABLE 2 The optimization problem and results of Study A.

Optimization problem			
Objective	Maximize Titer		
Constraints	$-0.5 \leq \text{CSFR}_0, \text{CSFR}_1 \leq 0.5$ $-1 \leq T, \text{pH} \leq 1$ $\text{Via} \geq 80\%$		
Results			
CSFR ₀	CSFR ₁	T	pH
0.5	0.5	1 (or 34°C)	1 (or 7.1)
Titer (prediction)	Via (prediction)		
$8.21 \pm 0.371 \text{ g/L}$	$(82.7 \pm 4.67) \%$		

discretization method. The response variables in Table 1 are the EoR HPLC titer (Titer) and the EoR viability (Via).

Based on the factor selection methodology described previously, we created the titer and viability models (i.e., Model A) as follows.

$$\text{Titer} = 6.39 + 0.53 \times \text{CSFR}_0 + 0.50 \times T + 0.47 \times \text{pH} + 0.58 \times \text{CSFR}_0 \times T + 0.60 \times \text{CSFR}_0 \times \text{pH} \left(R_{adj}^2 = 0.88 \right).$$

$$\text{Via} = 75.79 - 3.54 \times \text{CSFR}_0 + 5.66 \times \text{CSFR}_1 - 2.16 \times T + 6.00 \times \text{pH} + 3.24 \times \text{CSFR}_1 \times T - 2.74 \times \text{CSFR}_1 \times \text{pH} + 1.80 \times T \times \text{pH} \left(R_{adj}^2 = 0.84 \right).$$

The R_{adj}^2 value of each model is provided in parentheses. The lack of fit (LoF) p -values are 0.974 (not significant) for the titer model and 0.272 (not significant) for the viability model. An LoF that is not significant indicates that the corresponding model has represented all the non-random information in the data adequately.

In the titer model, all the main effects are positive—that is, the higher the CSFR, temperature, or pH, the higher the final titer. In addition, the average CSFR (CSFR_0) also has positive interactions with temperature and pH. In other words, when temperature and pH are higher, the positive effect from the average CSFR is stronger. Lastly, the effect of the slope of CSFR (CSFR_1) on the final titer is not significant hence not included in the titer model.

In the viability model, the effect of the average CSFR on viability is negative—while a higher CSFR is correlated with a higher titer, it is also correlated with lower viability. In addition, the slope of CSFR has a significant, strong, and positive effect on viability. This positive effect implies that a rising CSFR trend is associated with higher viability. Even though there are a few negative main and interaction effects in the model, the strongest among them are the positive effects from the slope of CSFR and the pH setpoint. Therefore, we can expect a rising CSFR trend and a higher pH to correlate strongly with higher viability.

The optimization problem and results are listed in Table 2. The optimization problem is to maximize the final titer while maintaining the final viability above 80%. $\text{CSFR}_0 = 0.5$, $T = 1$, and $\text{pH} = 1$ form the optimal condition that maximizes titer while satisfying the viability constraint. CSFR_1 is a free variable as it is not part of the titer model. That is to say, CSFR_1 can take any value within ± 0.5 without affecting

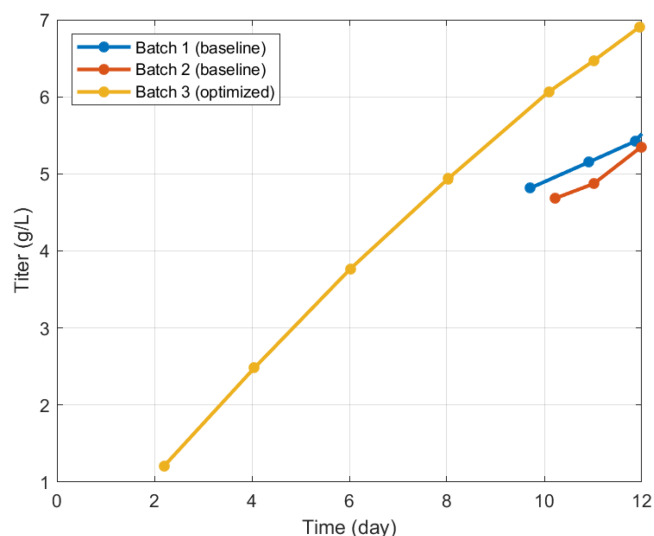


FIGURE 3 Comparison of titer from the baseline condition (Batch 1 and Batch 2) and the optimized condition (Batch 3).

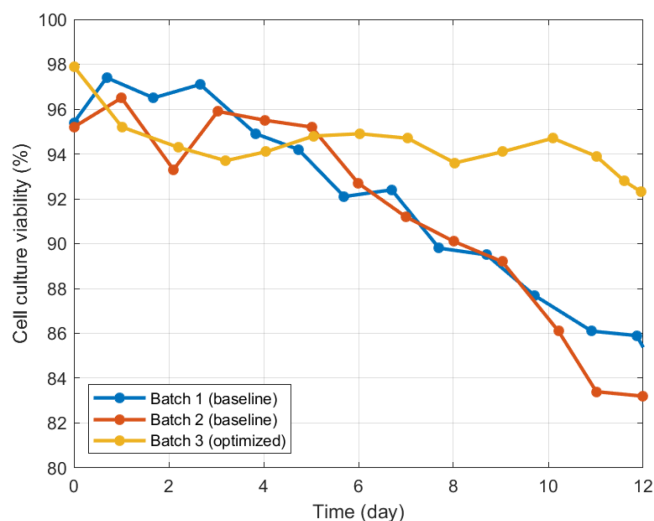


FIGURE 4 Comparison of viability from the baseline condition (Batch 1 and Batch 2) and the optimized condition (Batch 3).

the final titer. The final $CSFR_1$ was selected to be 0.5 as it both satisfies the viability constraint and maximizes viability. The 95% confidence intervals of the model-predicted responses were estimated and included. This optimized condition coincides with the condition used in Bioreactor/Run 18 in Table 1.

We tested the optimized condition in a 200-L process (Batch 3) and compared its titer and viability trends with two previous 200-L batches (Batch 1 and Batch 2) of the baseline process without dynamic feeding (Figures 3 and 4). Specifically, in Batch 1 and Batch 2, $CSFR$, temperature, and pH were kept at 0.0003 pL/cell/min, 33°C, and 6.95 respectively. The optimized condition in Batch 3 resulted in a 27% increase in the EoR titer, 3 days shorter to reach the previous EoR titer, and greater-than 92% viability at the end of the batch. This improvement in productivity and growth is unlikely due to batch-

to-batch variability as the differences in titer and viability between Batch 1 and Batch 2 are small compared with the differences between the baseline condition (Batch 1 and Batch 2) and the optimized condition (Batch 3). The model overestimated the EoR titer by 1.3 g/L. The model parameters were estimated using the data from small-scale Ambr250 batches, and the scale-to-scale variability could contribute to the observed bias between model prediction and measurement. In addition, the intrinsic complexity of cell culture processes introduces additional challenges in creating an accurate and generalizable model. Nevertheless, we can still attribute the substantial increase in titer primarily to the changes in $CSFR$, temperature, and pH.

We created a second DoDE to study any additional effects of the shape of a $CSFR$ trend on productivity. In this Ambr250 study (Study B), we used three DoDE factors— $CSFR_0$ and $CSFR_1$ in Study B are identical to $CSFR_0$ and $CSFR_1$ in Study A; $CSFR_2$ in the equation below represents the weight of the quadratic shifted Legendre polynomial, P_2 , in the $CSFR$ trend.

$$z(\tau) = CSFR_0 P_0(\tau) + CSFR_1 P_1(\tau) + CSFR_2 P_2(\tau) \\ = CSFR_0 + CSFR_1(-1 + 2\tau) + CSFR_2(1 - 6\tau + 6\tau^2),$$

where $CSFR_0$, $CSFR_1$, and $CSFR_2$ were scaled by the number of factors to prevent $CSFR$ from exceeding its bounds. The temperature and pH setpoints in Study B were identical to the optimized conditions from Study A (34°C and 7.1 pH). Each $CSFR$ trend was bounded between $CSFR_{\min} = 0.0002$ pL/cell/min and $CSFR_{\max} = 0.0005$ pL/cell/min. The $CSFR$ bounds were shifted such that the center point coincides with the optimal average $CSFR$ from Study A.

We formulated a 2^3 central composite design with six face-centered axial points and six center points in this study to capture any nonlinear effects (the design is orthogonal). We included axial points in Study B to study any higher-order effects or interactions that were not captured in the analysis of Study A. All the $CSFR$ trends are visualized in Figures 5 and 6. The number of discretization points was increased from one in Study A to 12 in Study B, that is, $CSFR$ changed once a day. The response variables in Table 3 are the EoR HPLC titer (Titer), the EoR viability (Via), and, additionally, the EoR N-linked manose-5 glycan (Man5). Man5 was selected as an additional response variable to illustrate any potential effect of feeding on antibody glycosylation as the glycoform is influenced by process conditions and can affect the in vivo function of an antibody.²¹

The models (i.e., Model B) are shown below.

$$\text{Titer} = 6.80 + 1.11 \times CSFR_0 - 0.44 \times CSFR_0^2 \left(R_{\text{adj}}^2 = 0.90 \right),$$

$$\text{Via} = 85.56 - 3.40 \times CSFR_0 + 3.51 \times CSFR_1 - 7.59 \\ \times CSFR_0^2 \left(R_{\text{adj}}^2 = 0.88 \right),$$

$$\text{Man5} = 4.28 + 0.99 \times CSFR_0 \left(R_{\text{adj}}^2 = 0.54 \right).$$

The LoF p -values are 0.747 (not significant) for the titer model, 0.0201 (significant) for the viability model, and 0.223 (not significant)

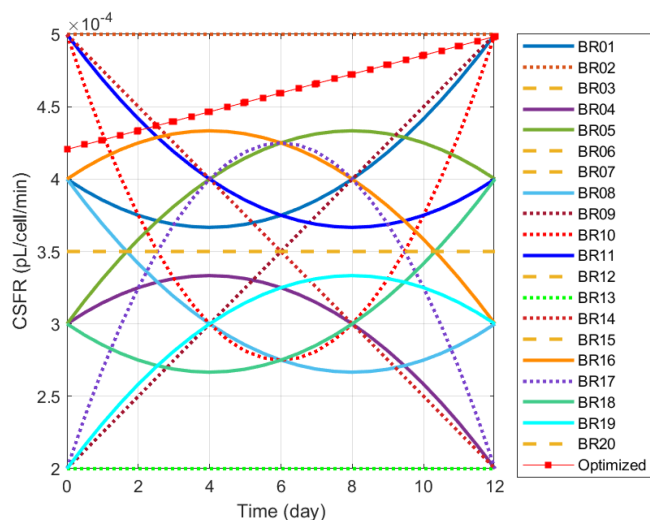


FIGURE 5 Theoretical CSFR trends in Study B. “BR” stands for “bioreactor” and is numbered according to the Run numbers in Table 3.

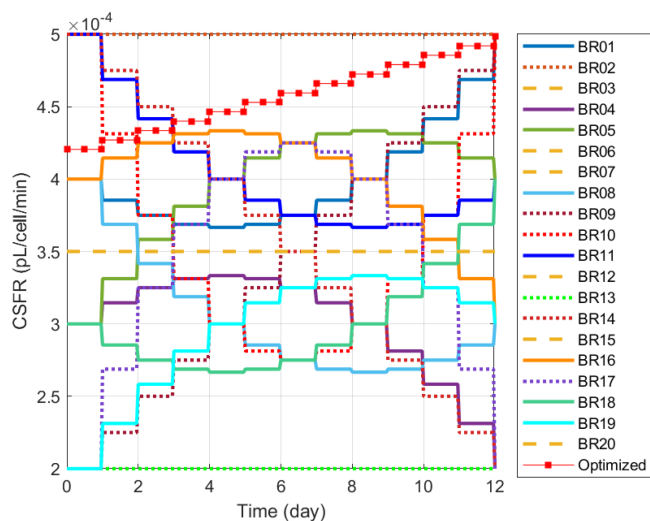


FIGURE 6 CSFR trends executed in Study B.

for the Man5 model. The significant LoF indicates that some non-random information in the viability data has not been represented by the model; however, we were not able to select another set of effect terms or a different factor transformation (e.g., logit transformation) that returned a not significant LoF. The R_{adj}^2 value of the Man5 model is low. We could improve the accuracy of the Man5 model to $R_{adj}^2 = 0.84$ by including $CSFR_1$, $CSFR_0 \times CSFR_1$, $CSFR_0^2$, $CSFR_1^2$, and $CSFR_0 \times CSFR_2^2$; however, the prediction R^2 would drop even lower from 0.23 to 0.051. We selected the Man5 model shown above as it is more “robust” even though it is less accurate than the high-order, more complex version. That said, the selection of the Man5 model does not affect the optimization results presented later.

The general observations from Study A still hold—that is, $CSFR_0$ has a positive effect on titer and a negative effect on viability; $CSFR_1$ does not have a significant effect on titer but has a positive effect on viability. We estimated several quadratic effects due to the inclusion of the axial points— $CSFR_0$ shows negative quadratic effects on both titer and viability. Lastly, $CSFR_0$ has a positive effect on Man5.

The third factor, $CSFR_2$, was found not to have any significant effect on titer, viability, or Man5. In other words, the zeroth-order ($CSFR_0$) and first-order ($CSFR_1$) properties of a CSFR trend contain sufficient variations associated with the responses; the marginal effect from any higher-order property is minimal. While it is still possible that quadratic or higher-order feed rate curves may be useful, describing feed rates with two DoDE factors is practical given the resource limitations.

The coefficients of Model A and Model B suggest a common trend that the average CSFR ($CSFR_0$) has a positive effect on titer and a negative effect on viability while the slope of CSFR ($CSFR_1$) has no significant effect on titer and a positive effect on viability. Specifically in this process, one needs to balance the opposite effects on titer and viability from $CSFR_0$ —under-feeding could lead to a lower yield, and over-feeding could lead to an early termination of the batch due to lower viability. To counter the negative effect on viability, one can feed the culture in increasing CSFRs to ensure viability above a certain value without affecting the final titer. The balancing of these trade-offs is reflected in the optimization problems in a quantitative and precise way by specifying numeric optimization objectives and constraints.

We formulated the following optimization problem to maximize titer while maintaining high viability and low Man5. We included the 5% Man5 upper bound as an additional constraint in Study B to illustrate the feasibility of maximizing productivity while meeting certain product quality requirements. The optimized CSFR in Table 4 is again a trend that increases linearly over time (like the optimized trend from Study A in Table 2). However, this optimized trend did not lead to a higher model-predicted titer. The maximized titer predicted by Model B is 0.84 g/L lower than the maximized predicted titer based on the results from Study A. In addition, the 95% confidence intervals of the two titer predictions do not overlap, implying that the difference is substantial (non-overlapping 95% confidence intervals of the means are sufficient to indicate that the difference is statistically significant given a significance level of 0.05 or lower but we did not estimate the actual p -value).

We evaluated the generalizability of the two models, which were created from the two experiments separately, by predicting Study B's responses using Model A. In other words, we tested Model A using an independent dataset. Factor levels of $CSFR_0$ and $CSFR_1$ in Study B were transformed such that they represent the equivalent levels defined in Study A, which has a different set of $CSFR_{min}$ and $CSFR_{max}$. Factor levels of T and pH were fixed at +1 as the temperature and pH levels were identical to the high levels in Study A. Study B's measurements and Model A's predictions of the EoR titer and viability are shown in Figures 7 and 8. We did not repeat the analysis in reverse and predict Study A's responses using Model B as Model B does not contain temperature or pH as inputs.

Run order	CSFR ₀	CSFR ₁	CSFR ₂	Titer (g/L)	Via (%)	Man5 (%)
1	0.33	0.33	0.33	7.07	85.66	5.13
2	1	0	0	7.44	73.96	4.40
3	0	0	0	6.89	85.15	4.46
4	-0.33	-0.33	-0.33	6.59	85.31	3.92
5	0.33	0.33	-0.33	7.34	85.04	5.10
6	0	0	0	7.01	86.23	4.63
7	0	0	0	6.79	86.47	4.46
8	-0.33	-0.33	0.33	6.16	82.16	3.64
9	0	1	0	6.59	87.24	4.23
10	0	0	1	6.84	85.72	4.44
11	0.33	-0.33	0.33	7.03	83.92	4.52
12	0	0	0	6.77	85.55	4.51
13	-1	0	0	5.27	81.74	3.01
14	0	-1	0	6.73	81.30	4.33
15	0	0	0	6.78	85.88	4.57
16	0.33	-0.33	-0.33	7.16	82.28	4.54
17	0	0	-1	6.98	84.09	4.53
18	-0.33	0.33	0.33	6.30	87.72	3.52
19	-0.33	0.33	-0.33	6.44	87.83	3.77
20	0	0	0	6.48	86.07	3.89

TABLE 3 The design and responses of Study B.

TABLE 4 The optimization problem and results of Study B.

Optimization problem		
Objective	Maximize Titer	
Constraints	$-1 \leq \text{CSFR}_0, \text{CSFR}_1, \text{CSFR}_2 \leq 1$ $ \text{CSFR}_0 + \text{CSFR}_1 + \text{CSFR}_2 \leq 1$ Via $\geq 80\%$ Man5 $\leq 5\%$	
Results		
CSFR ₀	CSFR ₁	CSFR ₂
0.73	0.26	0
Titer (prediction)	Via (prediction)	Man5 (prediction)
7.37 ± 0.185 g/L	(79.9 ± 1.38) %	(5.00 ± 0.359) %

Model A overestimated Study B's EoR titer by 1.48 g/L on average. On the other hand, the measured and predicted values are strongly correlated with Pearson $r^2 = 0.832$. This result suggests that Model A can explain Study B's titer qualitatively—the model can accurately predict whether a condition results in a higher (or lower) titer than another condition, but not the actual titer. One could potentially improve Model A's prediction accuracy by reducing the fixed effect. Such a bias resembles the over-prediction of the EoR titer in the 200-L testing experiment. Other contributions to the bias are the inherent batch-to-batch variability between Study A and Study B and a lower initial VCC used in Study B. In addition, CSFR was changed on Day 5 in Study A and changed daily in Study B due to different discretization methods used to implement the CSFR curves in the Ambr250 system. Such a difference could also potentially contribute

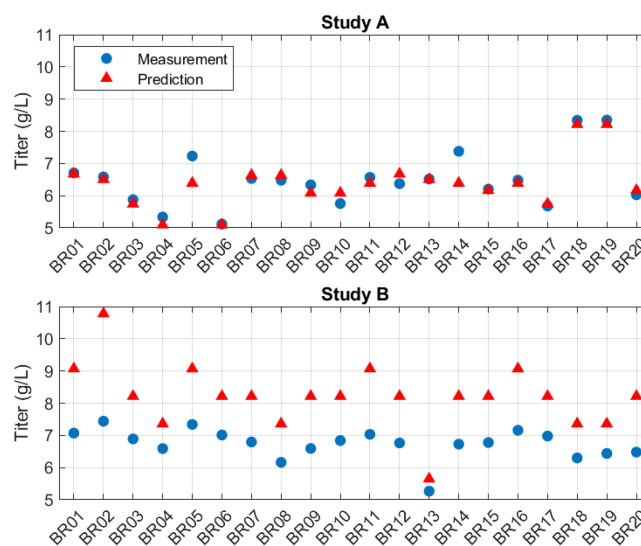


FIGURE 7 Measured and predicted (Model A) titer responses.

to the observed prediction errors. Nevertheless, the results imply that the strong effect of the average CSFR (CSFR₀) on titer remains unchanged in Study B. Model A underestimated viability in Study B by 4.81% on average (the measured and predicted values are moderately correlated with Pearson $r^2 = 0.349$).

The results above highlight the strengths and limitations of DoE/DoDE-based statistical models. On the one hand, it requires little additional knowledge about a process to develop a statistical model, and the model can achieve a high prediction accuracy if the experiment is designed appropriately with meaningful factors. That model

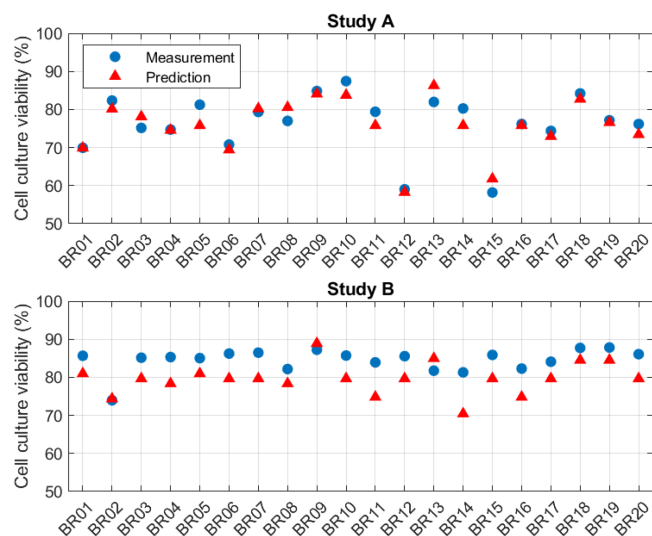


FIGURE 8 Measured and predicted (Model A) viability responses.

can then be used for optimizing process conditions as described in this article.

On the other hand, process dynamics, initial conditions, and other complex relationships are assumed to have no effect on the responses and are often omitted when creating a statistical model—sometimes, that assumption may not be valid, and a model based on it may be inaccurate. In addition, interpreting DoDE results might be more challenging than interpreting conventional DoE results due to the extra transformation between design factors and time trends. Avoiding higher-order Legendre polynomials when possible and exercising caution when executing the time-varying conditions are necessary for a DoDE study to be successful.

4 | CONCLUSION

In this article, we introduce the application of DoDE to cell culture process development, specifically the optimization of time-varying CSFR trends in fed-batch mAb manufacturing processes. We demonstrated the usefulness of this approach in two Ambr250 studies and verified the improvement in productivity experimentally in a large-scale experiment. DoDE can be implemented via widely-available, off-the-shelf software; therefore, the implementation will not incur additional process development cost. On the contrary, an efficient DoDE can potentially save process development cost compared with an iterative, OFAT design. We also found a common relationship between CSFR and process responses such as titer and viability—a CSFR trend with a higher average amount is associated with a higher titer but lower viability, and a CSFR trend that increases over time is associated with higher viability. We anticipate the application of DoDE in process development to be a starting point for more efficient, systematic, and standardized workflows across different assets in cell culture process development.

To replicate the DoDE-based optimization in this article, one needs to consider the following points. First, a conventional DoE with

static process conditions might be necessary prior to determining the DoDE factor levels. For instance, $CSFR_{min}$ and $CSFR_{max}$ in Study A were determined based on the data from the prior experiments such that cells would not be over-fed or experience severe nutrient depletion. Second, in the case of CSFR, two DoDE factors were found to be sufficient for titer optimization purposes. The marginal benefit of including more than two factors for CSFR was minimal. Third, one needs to balance between the depth and the breadth of a DoDE. For example, in Study A, we focused on testing a wide variety of factors and discovered the significant interactions among CSFR, pH, and temperature. However, due to the breadth of the design, we were not able to dive deeper into the nonlinear effects or the effects of the shape of CSFR on process responses.

The data from a DoDE study are also suitable for developing dynamic process models as the data contain variations in both process outputs and process inputs. Unlike statistical models such as the RSM models used in this article, a dynamic process model can be used to simulate the entire trajectory of multiple process variables of a batch instead of only the single instances of those process variables (e.g., the EoR titer). Using a dynamic process model, one can optimize the process inputs more flexibly and develop dynamic input profiles that are not constrained by the shapes of the basis polynomials. Furthermore, a dynamic process model used in conjunction with a state estimator and a model predictive controller can enable advanced process control of productivity and product quality attributes.⁷ The development of such a dynamic process model is not the scope of this article and will be explored in the future works.

AUTHOR CONTRIBUTIONS

Yu Luo: Conceptualization (lead); formal analysis (lead); writing – original draft (lead). **Duane A. Stanton:** Formal analysis (lead); writing – review and editing (equal). **Rachel C. Sharp:** Investigation (equal); writing – review and editing (equal). **Alexis J. Parrillo:** Investigation (equal); writing – review and editing (equal). **Kelsey T. Morgan:** Investigation (equal); writing – review and editing (equal). **Diana B. Ritz:** Writing – review and editing (equal). **Sameer Talwar:** Writing – review and editing (equal).

ACKNOWLEDGMENTS

The authors thank Tim Boung Wook Lee for the suggestion of using DoDE, John Craven, Tierney DeGennaro, Kevin Huff, and Jonathan Raley for their support and supervision of the Ambr250 and large-scale experiments, and Zach Hatzenbeller for checking the data used in this article.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at <https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/btpr.3380>.

DATA AVAILABILITY STATEMENT

No additional data are available for sharing.

ORCID

Yu Luo  <https://orcid.org/0000-0002-1379-6373>

Rachel C. Sharp  <https://orcid.org/0000-0002-6676-2249>

Alexis J. Parrillo  <https://orcid.org/0000-0002-7918-3764>

Kelsey T. Morgan  <https://orcid.org/0000-0003-2530-015X>

Sameer Talwar  <https://orcid.org/0000-0002-7670-0439>

REFERENCES

- Xu J, Xu X, Huang C, et al. Biomanufacturing evolution from conventional to intensified processes for productivity improvement: a case study. *mAbs*. 2020;12(1):1770669.
- Biegler LT, Grossmann IE. Retrospective on optimization. *Comp Chem Eng*. 2004;28(8):1169-1192.
- Štor J, Ruckerbauer DE, Szélová D, Zanghellini J, Borth N. Towards rational glyco-engineering in CHO: from data to predictive models. *Curr Opin Biotechnol*. 2021;71:9-17.
- Troup GM, Georgakis C. Process systems engineering tools in the pharmaceutical industry. *Comp Chem Eng*. 2013;51:157-171.
- Rathore AS, Nikita S, Thakur G, Mishra S. Artificial intelligence and machine learning applications in biopharmaceutical manufacturing. *Trends Biotechnol*. 2022;41(4):497-510.
- Klebanov N, Georgakis C. Dynamic response surface models: a data-driven approach for the analysis of time-varying process outputs. *Indus Eng Chem Res*. 2016;55(14):4022-4034.
- Luo Y, Kurian V, Ogunnaike BA. Bioprocess systems analysis, modeling, estimation, and control. *Curr Opin Chem Eng*. 2021;33:100705.
- Khuri AI, Mukhopadhyay S. Response surface methodology. *Wiley Interdiscip Rev Comput Stat*. 2010;2(2):128-149.
- Box GEP, Hunter JS, Hunter WG. *Statistics for Experimenters: Design, Innovation, and Discovery*. 2nd ed. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc.; 2005.
- Georgakis C. Design of Dynamic Experiments: a data-driven methodology for the optimization of time-varying processes. *Indus Eng Chem Res*. 2013;52(35):12369-12382.
- Castaldello C, Facco P, Bezzo F, Georgakis C, Barolo M. Data-driven tools for the optimization of a pharmaceutical process through its knowledge-driven model. *AIChE J*. 2022;69(4):e17925.
- Trentin G, Bertucco A, Georgakis C, Sforza E, Barbera E. Using the design of dynamic experiments to optimize photosynthetic cyanophycin production by *Synechocystis* sp. *J Indus Eng Chem*. 2023;117:386-393.
- Salim T, Chauhan G, Templeton N, Ling WLW. Using MVDA with stoichiometric balances to optimize amino acid concentrations in chemically defined CHO cell culture medium for improved culture performance. *Biotechnol Bioeng*. 2022;119(2):452-469.
- Lu F, Toh PC, Burnett I, et al. Automated dynamic fed-batch process and media optimization for high productivity cell culture process development. *Biotechnol Bioeng*. 2013;110(1):191-205.
- Eyster T, Talwar S, Fernandez J, et al. Tuning monoclonal antibody galactosylation using Raman spectroscopy-controlled lactic acid feeding. *Biotechnol Prog*. 2021;37(1):e3085.
- Gilbert A, McElearney K, Kshirsagar R, Sinacore MS, Ryll T. Investigation of metabolic variability observed in extended fed batch cell culture. *Biotechnol Prog*. 2013;29(6):1519-1527.
- Karra S, Sager B, Karim MN. Multi-scale modeling of heterogeneities in mammalian cell culture processes. *Indus Eng Chem Res*. 2010;49(17):7990-8006.
- Chusainow J, Yang YS, Yeo JHM, et al. A study of monoclonal antibody-producing CHO cell lines: what makes a stable high producer? *Biotechnol Bioeng*. 2009;102(4):1182-1196.
- Xie L, Wang DIC. High cell density and high monoclonal antibody production through medium design and rational control in a bioreactor. *Biotechnol Bioeng*. 1996;51(6):725-729.
- Oguchi S, Saito H, Tsukahara M, Tsumura H. pH condition in temperature shift cultivation enhances cell longevity and specific hMab productivity in CHO culture. *Cytotechnology*. 2006;52(3):199-207.
- Jefferis R. Glycosylation of recombinant antibody therapeutics. *Biotechnol Prog*. 2008;21(1):11-16.

How to cite this article: Luo Y, Stanton DA, Sharp RC, et al. Efficient optimization of time-varying inputs in a fed-batch cell culture process using design of dynamic experiments. *Biotechnol. Prog.* 2023;e3380. doi:10.1002/btpr.3380