RESEARCH

Optimizing Bioprocessing Efficiency with OptFed: Dynamic Nonlinear Modeling Improves Product-to-Biomass Yield by 19%

Guido Schlögel[1,2], Rüdiger Lück[3], Stefan Kittler[3], Oliver Spadiut[3], Julian Kopp[3], Jürgen Zanghellini[1] and Mathias Gotsmy[1,4*]

*Correspondence: mathias.gotsmy@univie.ac.at
[1]Department of Analytical Chemistry, University Vienna, Währinger Straße, 1090 Vienna, Austria, EU
[4]Austrian Centre of Industrial Biotechnology, Krenngasse 37, 8010 Graz, Austria, EU
Full list of author information is available at the end of the article

Abstract

Biotechnological production of a recombinant molecules relies heavily on fed-batch processes. However, as the cells' growth, substrate uptake, and production kinetics are often unclear, the fed-batches are frequently operated under sub-optimal conditions. For example, process designs are based on simple feed profiles (e.g., constant or exponential), operator experience, and basic statistical tools like response surface methodology (RSM), which are unable to harvest the full potential of the production processes.

To address this challenge, we propose a general modeling framework, OptFed, which utilizes experimental data from non-optimal fed-batch processes to predict an optimal process. In detail, we assume the cell-specific production rate depends on all state variables and their changes over time. Using measurements of bioreactor volume, biomass, and product, we train an ordinary differential equation model. To avoid overfitting, we use a regression model to reduce the number of kinetic parameters. Then, we predict the optimal process conditions (temperature and feed rate) by solving an optimal control problem using orthogonal collocation and nonlinear programming.

We apply OptFed to a recombinant protein L fed-batch production process. We determine optimal controls for feed rate and reactor temperature to maximize the product-to-biomass yield and successfully validate our predictions experimentally. Notably, our framework outperforms RSM in both simulation and experiments, capturing an optimum previously missed. We improve the experimental product-to-biomass ratio by 19 % and showcase OptFed's potential for enhancing process optimization in biotechnology.

Keywords: fed-batch; bioproduction; non-linear optimization; process design

1 Introduction

Biotechnological production processes are the backbone of numerous industries, from pharmaceuticals to biofuels. Many of these processes are operated as a fed-batch, adding substrate and nutrients continuously when the initial batch medium is depleted [31]. This method allows for control of key parameters, such as nutrient concentration, and is fundamental to achieving high yields and product quality. Consequently, the optimization of such processes becomes a critical objective.

Optimizing these processes can be challenging, given that the underlying mechanisms at the cellular level are complex and can vary significantly depending on factors such as product type, microorganism, induction mechanism, and product

location. As a result, optimization often involves a trial-and-error approach [45]. Alternatively, theoretical studies may employ simulations for the design of experiments [33, 36].

In general, there are two distinct paths for optimizing biotechnological processes: statistical design of experiments, such as Response Surface Methodology (RSM) [49], and the model-based approach [14, 27, 33, 36]. Statistical methods, including RSM, offer a straightforward and accessible means of optimization. However, they do not leverage biological knowledge, which could enhance the optimization process [34], and are limited to optimizing a predefined set of discrete variables [11]. In contrast, the model-based approach begins by understanding the underlying process, representing it accurately without relying on mathematical assumptions like the quadratic dependencies used in purely statistical methods [25].

For the model-based approach, empirical models are typically employed, often formulated as Ordinary Differential Equations (ODEs). For instance, Monod's widely used model for population growth [38] is one such example. Product creation in these models is often linked to growth rate [28, 32, 33] or feed rate [23]. These models provide a simple yet effective means of understanding and predicting bioprocess behavior, particularly when dealing with the often limited and noisy data typically obtained from bioprocesses.

However, simplicity comes at a cost. The inherent simplifications in these empirical models constrain their applicability, limiting their ability to describe complex processes comprehensively. These straightforward models often fail to account for some phenomena observed under constant process conditions on a cellular level. This includes production deterioration over time due to metabolic adaptation, or product inhibition [55]. Despite their limitations, these models are frequently used because formulating more complex models is hindered by sparse data, biological variation, and the difficulties in selecting the best available model equations [21]. Moreover, decisions in the modeling process can introduce biases that influence the outcome [18, 52], and using more complex models with small datasets may lead to overfitting [20].

The modeling process does not exist in isolation; its primary goal is to improve the efficiency of the process (e.g. by maximizing titer/biomass or minimizing operational costs). This objective often involves defining specific targets for each process, frequently utilizing the TRY metric (titer, rate, yield) [40]. The choice of optimization algorithms varies depending on the complexity of the model at hand, which is influenced by the available data set and its quality. While straightforward maximization algorithms suffice for discrete variables [41], optimizing continuous solutions, such as feed and temperature functions, poses mathematical challenges, notably due to the theoretically infinite number of control variables [24]. To address these challenges, various mathematical methods have been developed. For simpler models, Euler-Lagrange Equation-based approaches can be employed [33]. In contrast, more intricate models, as encountered in our work, require the application of optimal control theory [30].

The realm of optimal control problems has been extensively explored, resulting in a multitude of solution methods [48]. Analytical methods, such as those grounded in Pontryagin's Maximum Principle [30], are well-suited for relatively straightforward

problems but may not be practical for complex real-life scenarios. In most cases, numerical solvers become essential. The two primary categories of solvers are direct methods and those based on dynamic programming. Direct methods tackle the optimization and differential equations simultaneously, transforming the problem into a set of nonlinear differential equations [53]. In contrast, dynamic programming [8, 10] relies on the principle of optimality within Hamilton-Jacobi-Bellman frameworks, iteratively solving the problem. While dynamic programming holds the promise of identifying global optima, it tends to be slower than direct methods and is often infeasible for high-dimensional problems. As a result, direct methods find more frequent applications, particularly in engineering contexts, where a wealth of software packages is available for implementation [2, 7, 19, 58].

In this paper, we present OptFed, a comprehensive framework using an ODE (ordinary differential equations) model to describe a bioprocess. We acknowledge that we cannot select the best suited model without data. Therefore we start with a detailed model, remove terms and parameters that do not improve the model statistically significantly, and select a model small enough not to overfit the training data. Based on this model we leverage optimal control theory to identify optimal feed rates and temperatures. Feed rates remain unbound by predefined feeding strategies like exponential or constant feed.

2 Methods

2.1 Modeling Framework

Our modeling framework comprises three key components:

(I) define,
(II) fit, and
(III) optimize.

In the first stage, we establish a general process model capable of representing a wide range of biotechnological (fed-)batch production processes. The second stage fits the initial model to specific process data to adjust the model's parameters and simplifies the model. Finally, the fitted process model is used to optimize control variables, ultimately maximizing a freely selectable objective function. A graphical overview of the modeling framework is given in Figure 1.

A list of all parameters and their respective symbols and units is given in Supplementary Table S1.

2.1.1 Stage I: Define

We consider the production of recombinant proteins by *E. coli* in a fed-batch process, described by the following standard system of differential equations [31]:

$$\dot{X} = \mu X - \frac{f}{V} X, \qquad X(0) = X_{\mathrm{r}0}, \tag{1a}$$

$$\dot{P} = \pi X - \frac{f}{V} P, \qquad P(0) = 0, \tag{1b}$$

$$\dot{G} = -\gamma X + \frac{f}{V}(G_{\mathrm{f}} - G), \qquad G(0) = 0 \tag{1c}$$

$$\dot{V} = f, \qquad V(0) = V_0, \tag{1d}$$

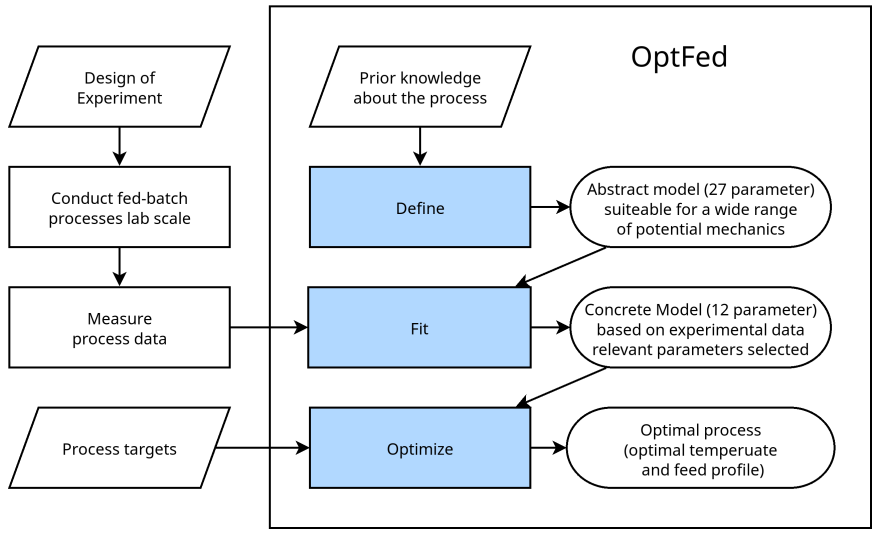


Figure 1: **Flow chart for OptFed**. In this study, we focus on the steps highlighted in blue; In the model definition we set up a general model including different inhibition effects. This model can describe different processes and uses many model parameters. In the second step, the model is simplified and parameters are fitted. This avoids overfitting and allows to fit a reasonable model with a moderate amount of data. Based on this model an optimal process is suggested.

where G, G_f, P, and X represent substrate concentrations in the reactor, substrate concentration in the feed, product, and total biomass, respectively, and V the current reactor volume. f, γ, μ, and π denote the feeding rate, substrate uptake rate per biomass, biomass growth rate, and specific product formation rate, respectively.

As the product is part of the biomass, we additionally define the metabolic active residual biomass (given as dry weight)

$$X_\mathrm{r} = X - P. \tag{1e}$$

This defines the uptake per active residual biomass (γ°) as

$$\gamma^\circ = \gamma X / X_\mathrm{r}. \tag{1f}$$

To connect the substrate uptake behavior with cellular growth and production, we assume that the total uptake can be divided into three additive components

$$\gamma^\circ = \gamma^\mu + \gamma^\pi + \gamma^\alpha, \tag{2}$$

where γ^μ, γ^π, and γ^α denote the specific substrate uptake supporting growth, product formation, and cellular maintenance, respectively. Here, maintenance summarizes all cellular processes not linked to growth or production.

Using biomass yield per substrate $Y_{\frac{X_\mathrm{r}}{G}}$ and product yield per substrate $Y_{\frac{P}{G}}$, we can connect glucose consumption to product formation and growth as follows:

$$\pi = \gamma^\pi Y_{\frac{P}{G}} X_\mathrm{r}/X, \quad \text{and} \quad \mu = \pi + \gamma^\mu Y_{\frac{X_\mathrm{r}}{G}} X_\mathrm{r}/X. \tag{3}$$

The factor X_r/X accounts for the fact that only the metabolically active residual biomass X_r contributes to additional product formation and growth.

We assume that total substrate uptake as well as substrate uptake for product formation follow a non-competitively inhibited Michaelis–Menten process [23, 35]

$$\gamma^\circ = \gamma^\circ_\mathrm{max}(T)\frac{G}{K^\circ_m + G}\prod_{i\in var_1}\frac{1}{1+i/K^\circ_i}, \quad var_1 = \{G, n, P/X, X\}, \tag{4a}$$

$$\gamma^\pi = \gamma^\pi_\mathrm{max}(T)\frac{\gamma^\circ - \gamma^\alpha}{K^\pi_m + \gamma^\circ - \gamma^\alpha}\prod_{i\in var_1}\frac{1}{1+i/K^\pi_i}, \tag{4b}$$

while substrate demand for maintenance is given by [51]

$$\gamma^\alpha = \gamma^\alpha_\mathrm{min}(T)\prod_{i\in var_2}(1+i/K^\alpha_i), \quad var_2 = \{\gamma^\circ, G, n, P/X, X\}. \tag{4c}$$

With these assumptions, γ^μ follows according to (2) to

$$\gamma^\mu = \gamma^\circ - \gamma^\pi - \gamma^\alpha. \tag{4d}$$

Here, K°_m, and K^π_m are Michaelis-Menten constants, while K°_i and K^π_i are inhibition constants, and K^α_i are activation constants. Finally, the minimum uptake rate $\gamma^\alpha_\mathrm{min}$, and the maximum uptake rates $\gamma^\circ_\mathrm{max}$ and γ^π_max are assumed to be temperature (T) dependent (excluding enzyme denaturation) [13]

$$\gamma^i_j(T) = E^i_0\frac{k_B T}{h}\frac{\exp\left(-\frac{\Delta G^i_\mathrm{cat}}{RT}\right)}{1+\exp\left[\frac{\Delta H^i_\mathrm{eq}}{R}\left(\frac{1}{T^i_\mathrm{eq}}-\frac{1}{T}\right)\right]}, \quad \begin{array}{l} i\in\{\mu,\pi,\circ\}, \\ j\in\{\mathrm{max},\mathrm{min}\}, \end{array} \tag{4e}$$

where E^i_0 is a (hypothetical) enzyme concentration, ΔG^i_cat, the activation energy, T^i_eq is the temperature where half of the enzymes are in an active state (half are inactivated by the high temperature), and ΔH^i_eq determines how abruptly the reaction rate declines with rising temperatures.

Our model, defined in equations (1) to (2), contains 29 free parameters. The substrate yields $Y_{\frac{X_\mathrm{r}}{G}}$, and $Y_{\frac{P}{G}}$ can be derived from genome-scale metabolic models [37], while the remaining 27 parameters need to be fitted from training data.

2.1.2 Stage II: Fit

Training Data Process data (biomass concentration, protein L concentration, and substrate concentration over time) from 12 fed-batch fermentations of *E. coli* strain BL21 DE3, with varying specific substrate feed and temperature [26], are used to train the general process model. To prevent overfitting, non-significant model terms are discarded (see below).

In short, protein L accumulated intracellularly, and glycerol was the sole carbon source. IPTG (isopropyl β-D-1-thiogalactopyranoside) was used to induce the product promoter during the feed phase. The biomass concentration at induction varied between $20\,\mathrm{g\,L^{-1}}$ to $45\,\mathrm{g\,L^{-1}}$. The control variables were the exponential feed rate

(f) and the temperature (T). Each process had a production phase of 12 h. Biomass and product concentrations were measured every 2 h, while temperature and feed rate were measured online. Due to the small reactor size, each sampling reduced the reactor volume non-negligibly, which was considered in the analysis. Details about the experimental setup and analytical methods are described in Section 2.4.1.

Experimental Data Interpolation and Rate Calculation The differential method [5, 17] is used to estimate uptake, growth, and production rates by fitting the concentration data and differentiating the fits. We fit splines, which are continuous in both their values and first derivatives, using SciPy's `UnivariateSpline` function [54] onto the experimental data points for control and state variables. By inserting these splines into equations (1) and (2), we calculate the experimental values for $\widehat{\gamma^{\circ}}$, $\widehat{\gamma^{\alpha}}$, $\widehat{\gamma^{\pi}}$, and $\widehat{\gamma^{\mu}}$ (see Supplementary Methods S1.1). The hat notation indicates that these variables are derived from the experimental training data and are used to estimate the unknown parameters in (4).

Model Parameter Estimation Parameter identification is performed using differential evolution [50] with SciPy's `differential_evolution` function [54], utilizing the previously calculated $\widehat{\gamma^{\circ}}$, $\widehat{\gamma^{\alpha}}$, and $\widehat{\gamma^{\pi}}$. Each of the three rates is fitted separately. The fitted model is then simplified by removing parameters that do not significantly reduce the fitting error. To determine significant error reduction, we adopted an algorithm inspired by ANOVA [3].

We assume that the effects of each variable in (4) are independent, meaning each effect is a separate factor. Each model factor, containing one influencing variable and one or more parameters, can be removed and the model remains valid. In case there is no temperature effect, equation (4e) simplifies to

$$\gamma_j^i(T) = c_{\gamma_j^i} \quad i \in \{\mu, \pi, \circ\},\ j \in \{\max, \min\}. \tag{5}$$

Each model factor is tested. If it does not improve the model fit statistically significantly, it is removed according to the following algorithm:

1. **Initial Fit**: Fit the model with all currently considered parameters by minimizing the sum of quadratic errors over all processes and measurement points. Calculate the error residuals and total variance for the measurement points. (Bounds used in the error minimization are shown in Supplementary Table S3.)
2. **Leave-One-Out Fit**: Repeat the fitting process for models, each missing one parameter.
3. **F-Test**: Use an F-test to determine if the reduction in variance is significant and calculate the difference in variance with and without the parameter.
4. **Remove Insignificant Parameters**: Remove the parameter with the highest p-value in the F-test.
5. **Iterate**: Repeat 1 to 4 until only significant terms remain[1].

Steps 1-5 are performed for each of the fitted rates ($\widehat{\gamma^{\circ}}$, $\widehat{\gamma^{\alpha}}$, and $\widehat{\gamma^{\pi}}$) separately.

[1]The significance level depends on the process design (e.g. number of measurement points). It is set with cross-validation data, selecting the significance level with the lowest error in the target variable (see Supplemental Methods S1.2).

2.1.3 Stage III: Optimize

To optimize

$$\max_{f(t),G_\mathrm{f},T(t),t_\mathrm{end}} \mathcal{Y}_{\frac{P}{X}} = \frac{P(t_\mathrm{end})}{X(t_\mathrm{end})} \tag{6a}$$

$$\text{s. t.} \quad \text{Eqn. (1)} \tag{6b}$$

$$V \leq V_\mathrm{max} = 2.5\,\mathrm{L} \tag{6c}$$

$$t_\mathrm{end} \leq t_\mathrm{max} = 12\,\mathrm{h} \tag{6d}$$

we use `ipopt` 3.14.10 [56], which is a general nonlinear programming solver. For discretization and numerical differentiation [see (1)], we implemented an orthogonal collocation and optimization algorithm in `casadi` 3.3.5 [2] using Python. All required code to reproduce our analysis is available at `https://github.com/gschloegel/OptFed`.

To solve the differential equations in (1), we first scale the time coordinate by setting $t = t_\mathrm{end}\tau$, using the process' time t_end as a control variable. We then apply orthogonal collocation on finite elements [9] to solve (1). Specifically, we use Gauss–Legendre polynomials of degree one with collocation points at 0.5. For the substrate, due to system stiffness, we use Gauss–Radau collocation points at 1. Locally, we solve the differential equation using the backward Euler method.

To avoid rapid variations in the control variables $u = (f, G_\mathrm{f}, T, t_\mathrm{end})$, specifically in the feed and temperature profile $f(t_\mathrm{end}\tau)$, and $T(t_\mathrm{end}\tau)$, we add a regularization term to the objective function in (6a). This modified objective reads

$$\max_{f(t),G_\mathrm{f},T(t),t_\mathrm{end}} \mathcal{Y}_{\frac{P}{X}} = \frac{P(t_\mathrm{end})}{X(t_\mathrm{end})} - \sum_{i=1}^{\mathrm{length}(u)} \frac{c_i}{h} \sum_{j\in\mathcal{S}} \left[\frac{u_i(j) - u_i(j+1)}{u_i(j) + u_i(j+1)}\right]^2, \tag{7}$$

where c_i is the penalty factor for each control variable u_i, h is the length of the finite elements, and $\mathcal{S}$ is the set of all sampling times.

In addition to the general optimization problem in (6), we defined a simplified version where all substrate is immediately used and no substrate accumulates, i.e. $\dot{G} = 0 = -\gamma^\circ X + \frac{f}{V}(G_\mathrm{f} - G)$. This mirrors the standard assumption in fed-batch processes and computationally avoids issues posed by stiff differential equations.

2.2 Response Surface Methodology

To benchmark OptFed, we evaluate its performance against response surface methodology (RSM) in process optimization. We calculated RSM using the `rsm` package [29] for R [43].

RSM requires that the control variables remain constant throughout the process and uses the process target metric (such as production concentration, yield, or productivity) to fit the model. Thus the RSM model is represented as:

$$\mathcal{Y}_{\frac{P}{X}} = c + c_f\mu_f + c_T T + c_{fT}\mu_f T + c_{f^2}\mu_f^2 + c_{T^2}T^2 \tag{8}$$

where c, c_f, c_T, c_{fT}, c_{f^2}, and c_{T^2} are fitted from the data minimizing the sum of quadratic errors. The same model is calculated with the target variables P and X.

2.3 Model Comparison and Validation

To evaluate model fit [Eqn. (1)], individual (state) variables (e.g. P, P/X, X) or specific rates (e.g. γ°, γ^α, and γ^π), we use the coefficient of determination R^2 and its adjusted version R^2_{adj}, which are defined as:

$$R^2 = 1 - \frac{RSS}{TSS} \qquad R^2_{\text{adj}} = 1 - \frac{RSS/df_{\text{res}}}{TSS/df_{\text{tot}}} \tag{9a}$$

with the residual and total sum of squares (RSS and TSS)

$$RSS = \sum_{p\in\mathcal{P}} \sum_{m\in\mathcal{S}} \left(x_p(m) - \widehat{x}_p(m)\right)^2 , \quad TSS = \sum_{p\in\mathcal{P}} \sum_{m\in\mathcal{S}} \left(x_p(m) - \langle\widehat{x}\rangle\right)^2 , \tag{9b}$$

respectively. Here, $x_p(m)$ and $\widehat{x}_p(m)$ represent the predicted and measured values respectively, $\langle\widehat{x}\rangle$ represents the average of the observed values, $\mathcal{P}$ is the set of all processes, and $\mathcal{S}$ is the set of all sampling times. df_{res} and df_{tot} are the residual and total degrees of freedom, given by $df_{\text{res}} = \#p - \#v - 1$ and $df_{\text{tot}} = \#p - 1$, respectively, where $\#p = |\mathcal{P}|+|\mathcal{S}|$ represents the number of points and $\#v$ represents the number of variables.

Finally, we measure relative errors with respect to the mean of all data points:

$$x_{\text{err}} = \frac{\widehat{x} - x}{\langle\widehat{x}\rangle} , \qquad \langle\widehat{x}\rangle = \frac{1}{|\mathcal{P}| \cdot |\mathcal{S}|} \sum_{p\in\mathcal{P}} \sum_{m\in\mathcal{S}} \widehat{x}. \tag{10}$$

Different models are compared using R^2 and R^2_{adj} on the state variables, focusing on P/X. In addition, we perform cross-validation using the leave-one-out strategy (predicting one process using all other processes) and compare R^2 for this as well.

To validate the results we do experimental validation, and test at least the predicted optimum against a new reference process (center point of the initial DoE).

2.4 Case study

We illustrate our modeling framework by optimizing protein L production in a fed-batch fermentation of *E. coli*, and evaluate its effectiveness in comparison to RSM [29] in central composite design [25]. Data from twelve fermentations with varying specific substrate feed and temperature [26] are used as training input for both methods to predict an optimal bioprocess that maximizes the specific process yield $\mathcal{Y}_{\frac{P}{X}}$.

The dataset represents nine conditions (see Figure 2a):

- one center point (four runs),
- four star points (single runs) where either feeding rate or temperature varied from the center point,
- four factorial points (single runs) where both variables deviated from the center point.

2.4.1 Experimental procedures

Experimental data for the training and validation of the proposed process optimization framework was obtained from 15 bioreactor cultivations. The cultivations were

carried out in two bioreactor systems having a similar working volume. The 12 initial cultivations were executed in a DASGIP© Parallel Bioreactor System (max. working volume 2 L; Eppendorf, Hamburg, Germany) as described in [26]. In contrast to the original paper, where 11 out of 12 performed processes were used, according to the DoE, we use all 12 processes as training data. In addition, three validation runs were performed in a Minifors 2 bioreactor system (max. working volume 2.5 L; Infors HT, Bottmingen, Switzerland). For all cultivations a defined minimal medium according to DeLisa [15] was used, supplemented with an initial concentration of 20 $g\,L^{-1}$ glycerol as the main carbon source and 0.02 $g\,L^{-1}$ kanamycin as a selection marker. The temperature was set to 37 °C during batch phase, 35 °C during fed-batch phase and controlled at defined levels during induced fed-batch phase in accordance with the experimental plan. The pH was monitored with an EasyFerm pH electrode (Hamilton, Reno, NV, USA) and kept constant at 6.7 via addition of 12.5% NH_4OH. A probe for monitoring dissolved oxygen (dO_2) was installed (Visiferm DO425, Hamilton, Reno, NV, USA). The dissolved oxygen in the cell broth was kept over 40% through continuous stirring (1400 rpm) and aeration of 2 vvm, if needed pure oxygen was added to the air flow. Furthermore, the off-gas was analyzed with respect to O_2 and CO_2 concentrations via a Bluevary sensor (BlueSens Gas analytics, Herten, Germany) for real-time monitoring of the metabolic activity of the cells. The process parameters were logged and controlled using the bioprocess management system eve© (Infors HT, Bottmingen, Switzerland). The expression of recombinant protein L was induced by a one-point addition of sterile Isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After addition of the inducer, samples were taken every two hours for further process and product analytics.

An *E. coli* BL21 (DE3) strain transformed with a pET-24a(+) plasmid was used for the cultivations (GenBank accession no. AAA67503). The plasmid carries the codon-optimized genes coding the 5B (binding) protein L with a C-terminal His_6-tag. The recombinant protein L is expressed intracellularly. The cells were harvested and subsequent analytics were done. All subsequent analytical steps were realised with samples of 35 mL cell broth each. The cell broth was centrifuged (10 min, 21 000 rpm, 4 °C) and the supernatant was separated from the cell pellet and aliquoted (1 mL) for anion exchange chromatography. Biomass concentration was quantified by dry cell weight (DCW) in triplicates. Therefore, the cell pellet was washed with saline (0.9 wt.% NaCl), centrifuged with the same settings and dried at 105 °C for 48 h. In addition, the biomass concentration was determined via optical density measurements at 600 nm wavelength (OD600) in triplicates. Residual glycerol and metabolites in the cell-free supernatant were analyzed by a high performance liquid chromatography (HPLC) system (UltiMate 3000; Thermo Fisher, Waltham, MA) equipped with an Aminex HPX-87 H column (Bio-Rad Laboratories, Hercules, CA, USA). HPLC standards with various concentrations of protein L (0.063-1.0 g/L), glycerol (0.781-50 g/L) and acetate (1-10 g/L) were prepared separately. A sample volume of 10 mL cell broth was centrifuged (15 min, 14 000 rpm, 4°C) and the separated cell pellet was re-suspended in 40 mL lysis buffer (10 mM EDTA, 100 mM Tris, pH 7.4) and homogenized subsequently (7 passages, 1200 bar; PandaPLUS, Gea AG, Germany). After centrifugation of the

crude cell lysate (20 min, 14 000 rpm, 4 °C), the supernatant was analyzed using a reversed-phase HPLC method for protein L quantification based on a PpL standard calibration curve. The UltiMate 3000 HPLC system was equipped with a BioResolve reversed-phase Polyphenyl column (Waters Corporation, MA, USA). Further information about the analytical procedures can be found in [26].

3 Results

We illustrate our modeling framework by optimizing protein L production in a fed-batch fermentation of *E. coli*, and evaluate its effectiveness in comparison to RSM [29] in central composite design [25]. Data from twelve fermentations with varying specific substrate feed and temperature [26] are used as training input for both methods to predict an optimal bioprocess that maximizes the specific process yield $\mathcal{Y}_{\frac{P}{X}}$ (product per biomass ratio at the end of the process).

3.1 Response surface methodology predicts limited optimization potential

RSM utilizes a second-degree polynomial model to forecast the specific yield, $\mathcal{Y}_{\frac{P}{X}}$ as a function of the specific feeding rate and temperature (Equation 8). It predicts an optimal specific yield of $0.16\,\mathrm{g\,g^{-1}}$ near the center point at $\mu_f = 0.12\,\mathrm{h^{-1}}$ and $T = 31\,°\mathrm{C}$ (Figure 2a). However, the predicted yield improvement is small (+1 %) yet uncertain (adjusted $R^2 = 0.19$; Figure 2b), and none of the model's parameters are statistically significant (at a 0.95 confidence level, Supplementary Table S4).

3.2 OptFed identifies significant optimization potential at high temperature

3.2.1 Model Simplification and Parameter Estimation

Initially, we mitigate measurement errors of the state variables by fitting them with cubic splines (Supplementary Figure S1). Based on these splines, we parameterize our general model using the algorithm described in Section 2.1.2. We found that only 12 out of 27 parameters of the general process model are statistically significant and required (at the optimal significance level of 0.2, Supplementary Figure S6), with only a minor drop in the explained variance (Table 2). The difference of R^2 and adjusted R^2 is higher for the full model as the degrees of freedom are higher. The effect is especially pronounced for γ° as only 3 out of 12 processes are used for estimation (for other processes the substrate concentration is below the limit of quantification). Further reducing the model would increase the fitting error by at least one-third (Table 1). Thus, with the parameter values listed in Table 1, the reduced model reads:

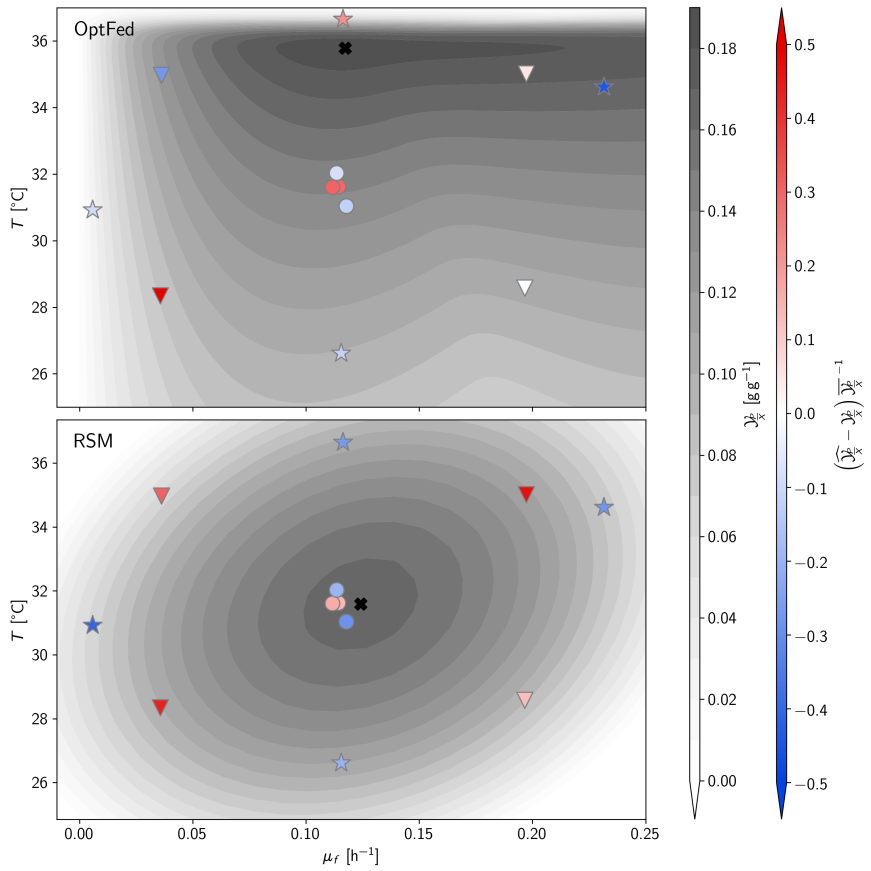
$$\gamma^\circ = c_{\gamma^\circ_{\max}} \frac{G}{K^\circ_m + G} \; \frac{K^\circ_G}{K^\circ_G + G} \tag{11a}$$

$$\gamma^\alpha = c_{\gamma^\alpha_{\min}} (1 + \gamma^\circ \; K^\alpha_g) \; (1 + P/X \; K^\alpha_P) \tag{11b}$$

$$\gamma^\pi = \gamma^\pi_{\max}(T) \frac{\gamma^\circ - \gamma^\alpha}{K^\pi_m + \gamma^\circ - \gamma^\alpha} \; \frac{1}{1 + n \; K^\pi_n} \tag{11c}$$

$$\gamma^\mu = \gamma^\circ - \gamma^\pi - \gamma^\alpha \tag{11d}$$

(a)



(b)

Variable	R^2(fit)		adjusted R^2(fit)		CV-R^2(cross validation)	
P/X	0.73	(0.56)	0.69	(0.19)	0.64	(-1.50)
P	0.82	(0.90)	0.79	(0.81)	0.75	(0.45)
X	0.83	(0.55)	0.81	(0.17)	0.66	(-1.14)

Figure 2: **Estimation of final product titer and model errors** (a) predicted specific yield in OptFed (top) versus RSM (bottom). Both models are constructed using data from twelve fermentations, including a center point (circle, four replicates), four factorial points (triangles), and four star points (stars; each single runs). The predicted specific product yields $\mathcal{Y}_{\frac{P}{X}}$ at the end of the process is shown in shades of gray, with the black cross indicating the optima. Both optima are calculated for exponential feed rates at constant temperature throughout the production phase. Model errors, indicating differences between predicted and measured $\mathcal{Y}_{\frac{P}{X}}$, are indicated by the color of the markers: red for overestimation, white for accurate prediction, and blue for underestimation. (b) Godness of fit [R^2(fit), and adjusted R^2(fit)] as well as goodness of fit for leave-one-out cross-validation (CV) [as measured by R^2(cross validation)]. Values in brackets refer to the RSM model, values in standard print to OptFed.

Table 1: List of parameters remaining in the selected model and their fitted values. Increased RSS next to a parameter shows the increase in fitting error (calculated with Equation 9) if this parameter would be removed from the reduced model. K_m° is not removed, as removal leads to physically impossible negative substrate concentrations. A comprehensive list of all parameters of the initially designed model is found in Supplementary Table S1).

	Name	Unit	Value	increased RSS
Parameters (fitted using training data)				
$c_{\gamma^\circ_{\text{max}}}$	maximal uptake rate	$\text{g}\,\text{g}^{-1}\,\text{h}^{-1}$	0.49	
$c_{\gamma^\alpha_{\text{min}}}$	maintenance requirement without growth	$\text{g}\,\text{g}^{-1}\,\text{h}^{-1}$	2.4×10^{-5}	
E_0^π	hypothetical enzyme concentration		9.8×10^{-9}	
K_m°	dependence on substrate concentration	g	1.0×10^{-3}	
K_g^α	growth dependent maintenance	$\text{g}\,\text{g}^{-1}\,\text{h}^{-1}$	1.0×10^{-4}	142 %
K_m^π	production dependence on available substrate	$\text{g}\,\text{g}^{-1}\,\text{h}^{-1}$	0.62	188 %
K_G°	substrate inhibition	$\text{g}\,\text{L}^{-1}$	89	119 %
K_P^α	increase caused by product	$\text{g}\,\text{L}^{-1}$	0.11	37 %
K_n^π	inhibition for higher no. of generations		1.5	50 %
$\Delta G_{\text{cat}}^\pi$	catalytic activation energy	$\text{J}\,\text{mol}^{-1}$	5.2×10^4	38 %
ΔH_{eq}^π	enthalpic (conversion of active to inactive enzyme)	$\text{J}\,\text{mol}^{-1}$	4.8×10^6	
T_{eq}^π	temperature where half the enzyme is active	K	310	
E_0^π	hypothetical enzyme concentration	K	9.8×10^{-9}	
Constants (from literature)				
$Y_{\frac{X_r}{G}}$	biomass yield per substrate	$\text{g}\,\text{g}^{-1}$	0.627	
$Y_{\frac{P}{G}}$	product yield per substrate	$\text{g}\,\text{g}^{-1}$	0.652	
k_B	Boltzmann constant	$\text{J}\,\text{K}^{-1}$	1.38×10^{-23}	
R	Gas constant	$\text{J}\,\text{mol}^{-1}\,\text{K}^{-1}$	8.314	
h	Plank constant	J h	2.39×10^{-30}	

Table 2: **Overview of the goodness of fit of the kinetic functions.** The values of the parameters of the reduced OptFed are given in Table 1.

	reduced OptFed			full OptFed		
	γ°	γ^α	γ^π	γ°	γ^α	γ^π
R^2	0.54	0.68	0.70	0.58	0.74	0.73
adjusted R^2	0.52	0.67	0.68	0.24	0.71	0.70
# of parameters	3	3	6	9	9	9

with

$$\gamma_{\max}^\pi(T) = E_0^\pi \frac{k_B T}{h} \frac{\exp\left(-\frac{\Delta G_{\text{cat}}^\pi}{RT}\right)}{1 + \exp\left(\frac{\Delta H_{\text{eq}}^\pi}{R}\left(T_{\text{eq}}^{\pi\,-1} - T^{-1}\right)\right)}. \tag{11e}$$

In the reduced OptFed process model (11), the substrate uptake rate γ° follows Michaelis-Menten kinetics with self-inhibition by the substrate. The substrate-to-maintenance flux γ^α increases multilinearly with the substrate uptake rate γ°, and the product-per-biomass ratio P/X. While both fluxes are temperature-independent, the product formation rate π is modeled as a temperature-dependent Michaelis-Menten-like kinetic with non-competitive inhibition by the number of generations n after induction. Supplementary Figure S3 illustrates the quality of our

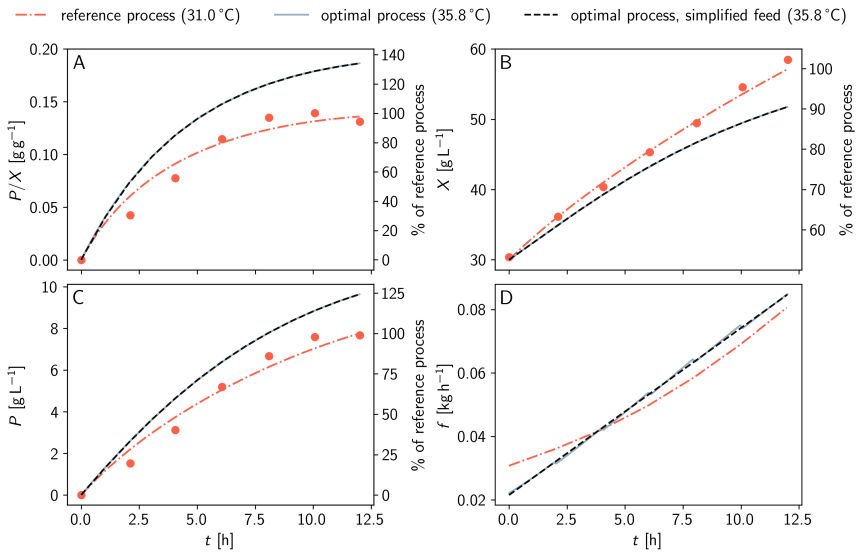


Figure 3: **Predicted optimal process with OptFed.** The red dashed-dotted lines represent the model fitted for the reference process, while grey lines show the predicted optimal behavior at $T = 35.8\,°\mathrm{C}$. black dashed lines indicate the optimal process with approximated linear feed rate. The simplification of the feed profile has no negative impact. After 12 h, a 37.4 % improvement in specific product yield is predicted. For reference, measurements from the training data are shown (orange circles, center point, same initial biomass).

model's fit on rates. Despite some noise and occasional large errors in individual data points, the overall trend is well-predicted.

We compare the measurement data for product and biomass with the model predictions and validate these predictions using cross-validation. The adjusted R^2 values remain above 0.52 for P, P/X, and X, both with and without cross-validation (Figure 2b and Supplementary Figures S4 and S5). Thus, we conclude that the model is a reliable choice, especially compared to RSM.

3.2.2 Process Model Optimization

Using the model develop above, we apply the optimization algorithm described in Methods Section 2.1.3 to determine the optimal feeding strategy and temperature settings. Additionally, we consider the following constraints:

- The initial biomass matches the mean value of the center point runs of the experimental data, $X_0 = 30\,\mathrm{g\,L^{-1}}$.
- The initial bioreactor volume is set at 1.3 L, with no maximum volume constraint.
- The feed glycerol concentration is fixed at $390\,\mathrm{g\,L^{-1}}$.
- Sampling of 35 mL reduces the current reactor volume at $t = 2, 4, 6, 8$, and 10 h.

- A linear regression model (Supplementary Figure S2) approximates the base addition as

$$f_{\text{base}} = XV \times 10^{-7} \times (13\,900\mu - 1.1\,\text{h}^{-1}).$$

Figure 3 illustrates the predicted optimal fed-batch process for protein L production. At 35.8 °C and after 12 h, we predict a biomass concentration of $52\,\text{g}\,\text{L}^{-1}$ and a product concentration of $9.7\,\text{g}\,\text{L}^{-1}$, resulting in an optimal product yield of $0.19\,\text{g}\,\text{g}^{-1}$. This represents a 37.4 % increase compared to the reference process.

Increasing the temperature is key for optimized performance (Figure 3). According to the model, maximum production rate $\gamma^{\pi}_{\text{max}}$ increases with temperature up to the optimal temperature of 35.8 °C and decreases sharply for higher temperatures (Supplementary Figure S3). Optimizing the feed but keeping the temperature at 31 °C increases the specific product yield by just 0.6 %. Conversely, keeping the feed constant and raising only the temperature boosts the maximally obtainable specific product yield by 37.2 %.

Unlike training fermentations, which used an exponential feed, we find that an almost linearly increasing feed rate is best to maximize the product-per-biomass yield (Figure 3D). Therefore, we decided to approximate the predicted optimal feed function with a simple linear equation

$$f_{\text{opt}} = 0.022\,\text{g}\,\text{h}^{-1} + 0.0053\,\text{g}\,\text{h}^{-2}\ t. \tag{12}$$

This adjustment changes the final product and biomass concentrations and the product-per-biomass-ratio ($\mathcal{Y}_{\frac{P}{X}}$) by less than 0.1 %, but significantly eases practical implementation (Figure 4). Generally, variations of the feed function have little influence as long as the initial and final biomass concentrations remain constant.

Our model's predictions are validated by experimentally running the optimal fermentation process with the simplified linear feed (Figure 4). Additionally, an intermediate process at 33.5 °C (halfway between the temperature of the center point and the optimum) is carried out.

For the first 6.5 h, the processes closely matched the predictions (Figure 4). The optimal process at 36 °C achieved a 19 % increase in specific product yield (compared to the predicted 21 %), while the process at 33.5 °C achieved a 5 % increase (compared to the predicted 18 %). Data points beyond 6.5 h are affected by unstable temperatures and therefore not considered.

4 Discussion

Recent years have seen a surge in research on biotechnological process optimization frameworks [6, 42, 44]. However, in many cases, the critical importance of the correct parameterization of process equations has not gained the deserved attention. Here, we developed a systematic modeling framework OptFed, to strategically address this blind spot.

OptFed employs a general phenomenological process model trained with experimental time series data from multiple fermentations. The algorithm discards parameters with insufficient statistical power to minimize overfitting, simplify the model,

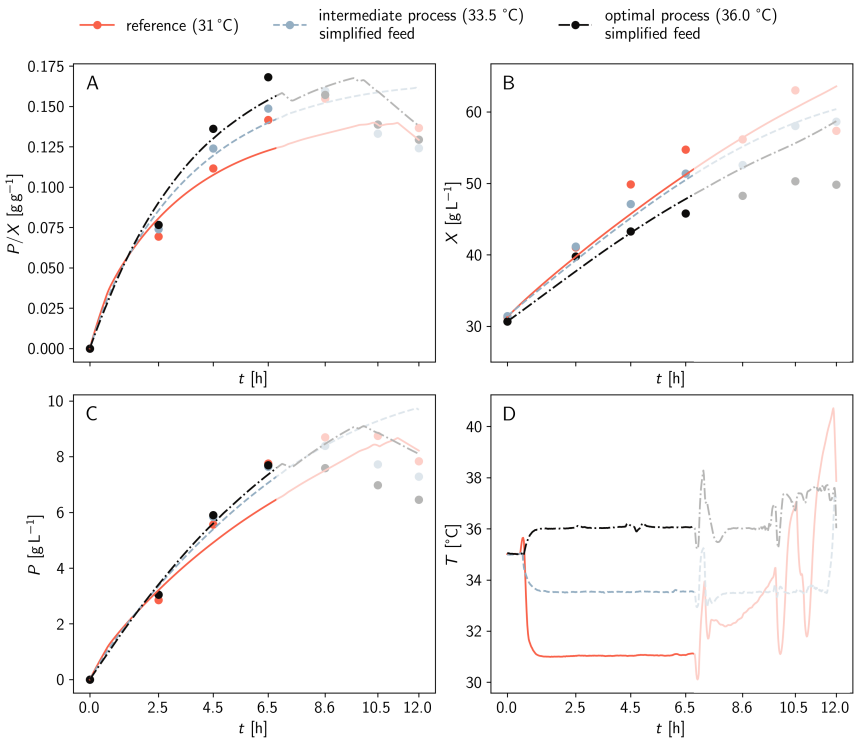


Figure 4: **Validation of the predicted optimal fermentation.** Panels A to C compare experimental data (circles) with model predictions (lines). Corresponding feed rates, substrate uptake rates, growth rates, and product formation rates are illustrated in Supplementary Figure S7. Panel D shows the temperature profile of the fermentations. For $t > 6.5$ h temperatures are unstable and the data is not considered. (opaque region in panels A to D).

and increase the model's reliability. Using this parameterized model, we applied nonlinear optimization to predict a fermentation profile for optimal specific protein L yield in *E. coli*, resulting in a near-linear feed at an elevated temperature of 35.8 °C. This approach predicted a 37 % increase in specific product yield compared to the training data, significantly outperforming the standard RSM, which only predicted a 1 % improvement. However, during the experimental validation of the optimized process, we encountered issues with temperature stability shortly after 6.5 h and decided not to use data collected thereafter. Despite this, at 6.5 h the specific protein yield was up by 19 %, close to the predicted 21 % at that time. The primary increase in specific product yield is attributed to the elevated temperature, accounting for over 99 % of the improvement. According to OptFed, product concentration increases approximately linearly with temperature up to the optimum. This dependence is completely missed by RSM, highlighting the advantage of our approach.

Compared to RSM, our process equations constrain the possible solution space to more realistic outcomes. In fact, combining mechanistic modeling with purely statistical approaches has already previously been shown to perform better than

pure RSM [42]. For example, RSM may predict negative values of the target metric (Figure 2), an effect that cannot be observed with OptFed. Moreover, our process equations implicitly ensure mass balance due to the calculation of the substrate-to-growth rate (γ^{μ}) from the difference of substrate uptake and the other substrate draining fluxes (Equation 2).

Maintenance flux, as we use it throughout the manuscript, is defined as a catchall-term for several metabolic effects. It comprises (1) the (non-)growth associated maintenance [12], (2) all non-optimal growth and production due to byproduct formation [4], and (3) any overflow metabolism during high substrate uptake rates [57]. Consequently, our yields are derived from a metabolomic model, excluding maintenance requirements, which differs from experimental yields where maintenance is included. Maintenance accounts for more than half of the uptake at the end of production (Supplementary Figure S10), leading to seemingly higher-than-usual yields ($Y_{\frac{X_r}{G}}, Y_{\frac{P}{G}}$) in Table 1.

Furthermore, we estimate uptake rates during production based on experimental data. As uptake can be significantly reduced during production [39] this avoids possible overfeeding in the predicted optimum. In the case study, we observe a reduced uptake rate, but uptake is not a limiting factor for optimization.

The mathematical problems in our procedure, such as model identification and calibration, are difficult to solve because models often have too many interrelated parameters and lack sufficient high-quality training data. This can cause optimization algorithms to be unstable and not converge, presenting significant challenges [1]. Our model, for example, requires high-quality time course data for training. Random (relative) errors for biomass and product concentrations should be in the range of 3 % and 15 %, respectively, to reliably identify the correct optimum (Supplementary Figure S8). However, these experimental uncertainties are manageable with current process monitoring technology [22].

Compared to the center point of the training data, almost all of the improvement in our case study originates from the increase in temperature. However, due to different product formation kinetics, this may be different for other products. Therefore, we cannot derive a general rule which feed profiles and temperatures are more advantageous in other setups. For example, [47] find a high influence of the feeding strategy on the production of inclusion bodies. This is most likely to a difference in product and process setup.

OptFed focuses on the model selection. Based on this further improvements are possible. Sensibility analyses [46] could help to make the optimum more stable considering the uncertainties in parameter fitting. We also limit ourselves by using existing data. Model-based design of experiment [16] could provide better training data or can be used to plan additional experiments to improve the model.

5 Conclusion

In this study, we present OptFed, a phenomenological model-based bioprocess optimization framework. In contrast to other tools, we emphasize the parameterization of process equations from training data, i.e., experimental values of bioprocesses. To avoid overfitting, OptFed provides a multi-step fitting strategy, where only terms that significantly decrease the model error are accepted whereas others are discarded. Further, we demonstrate that OptFed can not only describe the training

data but also predict enhanced process controls. Experimental validations show an increase in specific protein L yield by 19 % compared to the control.

We are confident that OptFed is a valuable tool for bioprocess optimization and will profit the industry in the future.

Competing interests
The authors declare that they have no competing interests.

Author's contributions
GS: conceptualization, methodology, software, investigation, visualization, writing – original draft, review and editing; RL: investigation, writing – review and editing; SK: investigation, writing – review and editing; OS: funding, writing – review and editing;, JK: investigation, writing – review and editing; JZ: funding, methodology, writing – original draft, review and editing; MG: methodology, writing – original draft, review and editing.

Author details
[1]Department of Analytical Chemistry, University Vienna, Währinger Straße, 1090 Vienna, Austria, EU. [2]Doctorate School of Chemistry, University of Vienna, Währinger Straße, 1090 Vienna, Austria, EU. [3]Integrated Bioprocess Development, Technical University Vienna, Getreidemarkt 9, 1060 Vienna, Austria, EU. [4]Austrian Centre of Industrial Biotechnology, Krenngasse 37, 8010 Graz, Austria, EU.

References
1. V. Abt, T. Barz, M. N. Cruz-Bournazou, C. Herwig, P. Kroll, J. Möller, R. Pörtner, and R. Schenkendorf. Model-based tools for optimal experiments in bioprocess engineering. *Current opinion in chemical engineering*, 22:244–252, 2018.
2. J. A. E. Andersson, J. Gillis, G. Horn, J. B. Rawlings, and M. Diehl. CasADi: a software framework for nonlinear optimization and optimal control. *Mathematical Programming Computation*, 11(1):1–36, Mar. 2019. ISSN 1867-2957. . URL `https://doi.org/10.1007/s12532-018-0139-4`.
3. Andrew Rutherford. *Introducing Anova and Ancova : A GLM Approach*. Introducing Statistical Methods. SAGE Publications Ltd, London, 2001. ISBN 978-0-7619-5160-5. URL `https://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=251737&site=ehost-live`.
4. A. A. Aristidou, K.-Y. San, and G. N. Bennett. Improvement of biomass yield and recombinant gene expression in escherichia coli by using fructose as the primary carbon source. *Biotechnology progress*, 15(1):140–145, 1999.
5. A. Bardow and W. Marquardt. Incremental and simultaneous identification of reaction kinetics: methods and comparison. *Chemical Engineering Science*, 59(13):2673–2684, July 2004. ISSN 0009-2509. . URL `https://www.sciencedirect.com/science/article/pii/S0009250904002015`.
6. J. Bauer and S. Klamt. Optmsp: A toolbox for designing optimal multi-stage (bio) processes. *Journal of Biotechnology*, 383:94–102, 2024.
7. L. D. R. Beal, D. C. Hill, R. A. Martin, and J. D. Hedengren. GEKKO Optimization Suite. *Processes*, 6(8):106, Aug. 2018. ISSN 2227-9717. . URL `https://www.mdpi.com/2227-9717/6/8/106`. Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.
8. R. E. Bellman. *Dynamic Programming*. Princeton University Press, Aug. 2021. ISBN 978-1-4008-3538-6. . URL `http://www.degruyter.com/document/doi/10.1515/9781400835386/html`. Publication Title: Dynamic Programming.
9. L. T. Biegler. *Nonlinear Programming: Concepts, Algorithms, and Applications to Chemical Processes*. Society for Industrial and Applied Mathematics, Jan. 2010. ISBN 978-0-89871-702-0 978-0-89871-938-3. . URL `http://epubs.siam.org/doi/book/10.1137/1.9780898719383`.
10. B. Bojkov and R. Luus. Time-Optimal Control by Iterative Dynamic Programming. *Industrial & Engineering Chemistry Research*, 33(6):1486–1492, June 1994. ISSN 0888-5885. . URL `https://doi.org/10.1021/ie00030a008`. Publisher: American Chemical Society.
11. J. C. M. Carvalho, M. Vitolo, S. Sato, and E. Aquarone. Ethanol production by Saccharomyces cerevisiae grown in sugarcane blackstrap molasses through a fed-batch process. *Applied Biochemistry and Biotechnology*, 110(3):151–164, Sept. 2003. ISSN 1559-0291. . URL `https://doi.org/10.1385/ABAB:110:3:151`.
12. B. L. Coltman, C. Rebnegger, B. Gasser, and J. Zanghellini. Characterising the metabolic rewiring of extremely slow growing komagataella phaffii. *Microbial Biotechnology*, 17(1):e14386, 2024.
13. R. M. Daniel, M. J. Danson, R. Eisenthal, C. K. Lee, and M. E. Peterson. The effect of temperature on enzyme activity: new insights and their implications. *Extremophiles*, 12(1):51–59, Jan. 2008. ISSN 1433-4909. . URL `https://doi.org/10.1007/s00792-007-0089-7`.
14. R. D. de Oliveira, G. A. Le Roux, and R. Mahadevan. Nonlinear programming reformulation of dynamic flux balance analysis models. *Computers & Chemical Engineering*, 170:108101, 2023.
15. M. P. DeLisa, J. Li, G. Rao, W. A. Weigand, and W. E. Bentley. Monitoring GFP-operon fusion protein expression during high cell density cultivation of Escherichia coli using an on-line optical sensor. *Biotechnology and Bioengineering*, 65(1):54–64, 1999. .
16. G. Franceschini and S. Macchietto. Model-based design of experiments for parameter precision: State of the art. *Chemical Engineering Science*, 63(19):4846–4872, Oct. 2008. ISSN 0009-2509. . URL `https://www.sciencedirect.com/science/article/pii/S0009250907008871`.
17. G. F. Froment, K. B. Bischoff, and J. De Wilde. *Chemical reactor analysis and design*, volume 2. Wiley New York, 1990.
18. G. Gigerenzer and H. Brighton. Homo Heuristicus: Why Biased Minds Make Better Inferences. *Topics in Cognitive Science*, 1(1):107–143, 2009. ISSN 1756-8765. . URL

https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-8765.2008.01006.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-8765.2008.01006.x.

19. W. E. Hart, J.-P. Watson, and D. L. Woodruff. Pyomo: modeling and solving mathematical programs in Python. *Mathematical Programming Computation*, 3(3):219, Aug. 2011. ISSN 1867-2957. . URL https://doi.org/10.1007/s12532-011-0026-8.
20. D. M. Hawkins. The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, 44(1): 1–12, Jan. 2004. ISSN 0095-2338. . URL https://doi.org/10.1021/ci0342472. Publisher: American Chemical Society.
21. H. W. Jannasch and T. Egli. Microbial growth kinetics: a historical perspective. *Antonie van Leeuwenhoek*, 63 (3):213–224, Sept. 1993. ISSN 1572-9699. . URL https://doi.org/10.1007/BF00871219.
22. J. Kager and C. Herwig. Monte Carlo-Based Error Propagation for a More Reliable Regression Analysis across Specific Rates in Bioprocesses. *Bioengineering*, 8(11):160, Nov. 2021. ISSN 2306-5354. . URL https://www.mdpi.com/2306-5354/8/11/160. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute.
23. J. Kager, J. Bartlechner, C. Herwig, and S. Jakubek. Direct control of recombinant protein production rates in E. coli fed-batch processes by nonlinear feedback linearization. *Chemical Engineering Research and Design*, 182: 290–304, June 2022. ISSN 0263-8762. . URL https://www.sciencedirect.com/science/article/pii/S0263876222001460.
24. D. Kalise, K. Kunisch, and Z. Rao. *Hamilton-Jacobi-Bellman Equations: Numerical Methods and Applications in Optimal Control*. De Gruyter, Aug. 2018. ISBN 978-3-11-054359-9. . URL http://www.degruyter.com/document/doi/10.1515/9783110543599/html. Publication Title: Hamilton-Jacobi-Bellman Equations.
25. A. I. Khuri and S. Mukhopadhyay. Response surface methodology. *WIREs Computational Statistics*, 2(2): 128–149, 2010. ISSN 1939-0068. . URL http://onlinelibrary.wiley.com/doi/abs/10.1002/wics.73. _eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wics.73.
26. S. Kittler, J. Ebner, M. Besleaga, J. Larsbrink, B. Darnhofer, R. Birner-Gruenberger, S. Schobesberger, C. K. Akhgar, A. Schwaighofer, B. Lendl, and O. Spadiut. Recombinant Protein L: Production, Purification and Characterization of a Universal Binding Ligand. *Journal of Biotechnology*, 359:108–115, Nov. 2022. ISSN 0168-1656. . URL https://www.sciencedirect.com/science/article/pii/S0168165622002371.
27. S. Klamt, R. Mahadevan, and O. Hädicke. When do two-stage processes outperform one-stage processes? *Biotechnology journal*, 13(2):1700539, 2018.
28. S. Klumpp, Z. Zhang, and T. Hwa. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*, 139(7):1366–1375, Dec. 2009. ISSN 0092-8674, 1097-4172. . URL https://www.cell.com/cell/abstract/S0092-8674(09)01505-0. Publisher: Elsevier.
29. R. V. Lenth. Response-Surface Methods in R, Using rsm. *Journal of Statistical Software*, 32:1–17, 2010. ISSN 1548-7660. . URL https://doi.org/10.18637/jss.v032.i07.
30. D. Liberzon. *Calculus of Variations and Optimal Control Theory*. Princeton University Press, 2012. ISBN 978-0-691-15187-8.
31. H. C. Lim and H. S. Shin. *Fed-batch cultures: principles and applications of semi-batch bioreactors*. Cambridge University Press, 2013.
32. M. B. Lopes, G. Martins, and C. R. C. Calado. Kinetic modeling of plasmid bioproduction in Escherichia coli DH5α cultures over different carbon-source compositions. *Journal of Biotechnology*, 186:38–48, Sept. 2014. ISSN 0168-1656. . URL https://www.sciencedirect.com/science/article/pii/S0168165614003137.
33. M. Maurer, M. Kühleitner, B. Gasser, and D. Mattanovich. Versatile modeling and optimization of fed batch processes for the production of secreted heterologous proteins with pichia pastoris. *Microbial cell factories*, 5 (1):1–10, 2006.
34. G. Mermoud. Model-Based Optimization. In G. Mermoud, editor, *Stochastic Reactive Distributed Robotic Systems: Design, Modeling and Optimization*, Springer Tracts in Advanced Robotics, pages 175–179. Springer International Publishing, Cham, 2014. ISBN 978-3-319-02609-1. . URL https://doi.org/10.1007/978-3-319-02609-1_11.
35. L. Michaelis and M. L. Menten. Die Kinetik der Invertinwirkung. *Biochemische Zeitschrift*, 49:333–369, 1913. ISSN 0366-0753. URL http://publikationen.ub.uni-frankfurt.de/frontdoor/index/index/docId/17273.
36. J. Modak, H. Lim, and Y. Tayeb. General characteristics of optimal feed rate profiles for various fed-batch fermentation processes. *Biotechnology and bioengineering*, 28(9):1396–1407, 1986.
37. J. M. Monk, A. Koza, M. A. Campodonico, D. Machado, J. M. Seoane, B. O. Palsson, M. J. Herrgård, and A. M. Feist. Multi-omics quantification of species variation of escherichia coli links molecular features with strain phenotypes. *Cell systems*, 3(3):238–251, 2016.
38. J. Monod. The growth of bacterial cultures. *Annual Review of Microbiology*, 3(1):371–394, Oct. 1949. ISSN 0066-4227. . URL https://www.annualreviews.org/doi/10.1146/annurev.mi.03.100149.002103. Publisher: Annual Reviews.
39. D. F. Müller, D. Wibbing, C. Herwig, and J. Kager. Simultaneous real-time estimation of maximum substrate uptake capacity and yield coefficient in induced microbial cultures. *Computers & Chemical Engineering*, 173: 108203, May 2023. ISSN 0098-1354. . URL https://www.sciencedirect.com/science/article/pii/S0098135423000728. Publisher: Pergamon.
40. J. Nielsen and J. D. Keasling. Engineering cellular metabolism. *Cell*, 164(6):1185–1197, 2016.
41. J. Nocedal and S. J. Wright. *Numerical optimization*. Springer series in operations research. Springer, New York, 2nd ed edition, 2006. ISBN 978-0-387-30303-1. OCLC: ocm68629100.
42. J. Pinto, C. R. de Azevedo, R. Oliveira, and M. von Stosch. A bootstrap-aggregated hybrid semi-parametric modeling framework for bioprocess development. *Bioprocess and biosystems engineering*, 42:1853–1865, 2019.
43. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2023. URL https://www.R-project.org/.

44. K. Raj, N. Venayak, and R. Mahadevan. Novel two-stage processes for optimal chemical production in microbes. *Metabolic Engineering*, 62:186–197, Nov. 2020. ISSN 1096-7176. . URL https://www.sciencedirect.com/science/article/pii/S1096717620301269.
45. M. I. Rodrigues and A. F. Iemma. *Experimental Design and Process Optimization*. CRC Press, Boca Raton, Dec. 2014. ISBN 978-0-429-16186-5. .
46. R. Schenkendorf, X. Xie, M. Rehbein, S. Scholl, and U. Krewer. The Impact of Global Sensitivities and Design Measures in Model-Based Optimal Experimental Design. *Processes*, 6(4):27, Apr. 2018. ISSN 2227-9717. . URL https://www.mdpi.com/2227-9717/6/4/27. Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
47. C. Slouka, J. Kopp, D. Strohmer, J. Kager, O. Spadiut, and C. Herwig. Monitoring and control strategies for inclusion body production in *E. coli* based on glycerol consumption. *Journal of Biotechnology*, 296:75–82, Apr. 2019. ISSN 0168-1656. . URL https://www.sciencedirect.com/science/article/pii/S0168165619300951.
48. B. Srinivasan, S. Palanki, and D. Bonvin. Dynamic optimization of batch processes: I. Characterization of the nominal solution. *Computers & Chemical Engineering*, 27(1):1–26, Jan. 2003. ISSN 0098-1354. . URL https://www.sciencedirect.com/science/article/pii/S0098135402001163.
49. D. M. Steinberg and D. Bursztyn. Response Surface Methodology in Biotechnology. *Quality Engineering*, 22 (2):78–87, Mar. 2010. ISSN 0898-2112. . URL https://doi.org/10.1080/08982110903510388. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/08982110903510388.
50. R. Storn and K. Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, 11(4):341–359, Dec. 1997. ISSN 1573-2916. . URL https://doi.org/10.1023/A:1008202821328.
51. P. van Bodegom. Microbial Maintenance: A Critical Review on Its Quantification. *Microbial Ecology*, 53(4): 513–523, May 2007. ISSN 1432-184X. . URL https://doi.org/10.1007/s00248-006-9049-5.
52. S. Varma and R. Simon. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics*, 7(1):91, Feb. 2006. ISSN 1471-2105. . URL https://doi.org/10.1186/1471-2105-7-91.
53. V. S. Vassiliadis, R. W. H. Sargent, and C. C. Pantelides. Solution of a Class of Multistage Dynamic Optimization Problems. 1. Problems without Path Constraints. *Industrial & Engineering Chemistry Research*, 33(9):2111–2122, Sept. 1994. ISSN 0888-5885. . URL https://doi.org/10.1021/ie00033a014. Publisher: American Chemical Society.
54. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020. .
55. J. Weber, F. Hoffmann, and U. Rinas. Metabolic adaptation of Escherichia coli during temperature-induced recombinant protein production: 2. Redirection of metabolic fluxes. *Biotechnology and Bioengineering*, 80(3): 320–330, 2002. ISSN 1097-0290. . URL https://onlinelibrary.wiley.com/doi/abs/10.1002/bit.10380. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.10380.
56. A. Wächter and L. T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, Mar. 2006. ISSN 1436-4646. . URL https://doi.org/10.1007/s10107-004-0559-y.
57. B. Xu, M. Jahic, and S.-O. Enfors. Modeling of overflow metabolism in batch and fed-batch cultures of escherichiacoli. *Biotechnology progress*, 15(1):81–90, 1999.
58. J. Yang, Y. Yang, and M. Li. OptControl.jl: An Interpreter for Optimal Control Problem, July 2022. URL http://arxiv.org/abs/2207.13229. arXiv:2207.13229 [math].