

RESEARCH

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

Guido Schlögel^{1,2}, Rüdiger Lück³, Stefan Kittler³, Oliver Spadiut³, Julian Kopp³, Jürgen Zanghellini¹ and Mathias Gotsmy^{1,4*}

*Correspondence:

mathias.gotsmy@univie.ac.at

¹Department of Analytical Chemistry, University Vienna, Währinger Straße, 1090 Vienna, Austria, EU

⁴Austrian Centre of Industrial Biotechnology, Krenngasse 37, 8010 Graz, Austria, EU

Full list of author information is available at the end of the article

S1 Supplementary Methods

S1.1 Spline fit

In Section 2.1.2, the fitting of the rates g , γ^μ , and γ^π is described. Here we go into more detail about how they are calculated for our specific experimental setup.

The following variables are measured during the processes of the training data (see Section 2.4.1):

T	Temperature
\hat{V}	Volume (as reactor mass)
f_{cum}^G	cumulative substrate feed (mass of the feed flask)
f_{cum}^B	cumulative base feed for control pH (mass of the feed flask)
G_f^G	feed concentration in the substrate feed
\hat{X}	biomass concentration
\hat{P}	product concentration
\hat{G}	free substrate in the media

The first 4 values are measured continuously, and the last 3 every 2 hours.

To get continuous functions for the biomass, product, and substrate concentrations and derivatives for the cumulative feeds, spline fitting is used:

$$f = f_{\text{cum}}^{\dot{G}} + f_{\text{cum}}^{\dot{B}} \quad (\text{S1a})$$

$$G_f = f_{\text{cum}}^{\dot{G}} G_f^G f^{-1} \quad (\text{S1b})$$

$$\hat{\mu} = \dot{\hat{X}} \hat{X}^{-1} + \frac{f}{\hat{V}} \quad (\text{S1c})$$

$$\hat{\pi} = \left(\dot{\hat{P}} + \frac{f}{\hat{V}} \hat{P} \right) \hat{X}^{-1} \quad (\text{S1d})$$

$$\hat{\gamma} = \left(\frac{f}{\hat{V}} (G_f - \hat{G}) - \dot{\hat{G}} \right) \hat{X}^{-1} \quad (\text{S1e})$$

$$\hat{\gamma}^{\circ} = \hat{\gamma} \left(1 - \frac{\hat{P}}{\hat{X}} \right)^{-1} \quad (\text{S1f})$$

$$\hat{\gamma}^{\pi} = \frac{\hat{\pi}}{Y_{\frac{P}{G}}} \left(1 - \frac{\hat{P}}{\hat{X}} \right)^{-1} \quad (\text{S1g})$$

$$\hat{\gamma}^{\mu} = \left(\hat{\mu} - \hat{\pi} Y_{\frac{P}{G}} \right) Y_{\frac{X_r}{G}}^{-1} \left(1 - \frac{\hat{P}}{\hat{X}} \right)^{-1} \quad (\text{S1h})$$

$$\hat{\gamma}^{\alpha} = \hat{\gamma}^{\circ} - \hat{\gamma}^{\mu} - \hat{\gamma}^{\pi} \quad (\text{S1i})$$

$\hat{\gamma}^{\circ}$ can only be estimated if the substrate concentration is above the limit of quantification.

To negate the effect of sampling we transform the feed and the volume by the sampling factor s . This creates an equivalent process without sampling where all concentrations (\hat{X} , \hat{P} , \hat{G}) and the specific feed (f/\hat{V}) are equal.

$$s(t) = \prod_{t_i \in S, t_i < t} \left(1 - \frac{V_{\text{sample}}}{\hat{V}(t_i)} \right) \quad (\text{S2a})$$

$$V_{\text{corr}}(t) = \frac{\hat{V}(t)}{s(t)} \quad (\text{S2b})$$

$$f_{\text{corr}}(t) = \frac{\hat{f}(t)}{s(t)} \quad (\text{S2c})$$

S is the set of all sample times and V_{sample} the sampling volume. Fitting and optimization are done for this corrected, unsampled process, and the results are transformed back to the sampled process.

S1.2 Cross validation and hyperparameter selection

In Section 2.1.2 the significance level α is treated as a hyperparameter and has to be defined by cross-validation, e.g., it depends on the sampling regime, more sampling points will require a lower α . For cross-validation, the leave-one-out strategy estimates each run in the training data with the 11 other training runs. For these prediction, R^2 is calculated according to (12).

We repeat the model fit and cross-validation for different levels of significance (0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, 0.4) and

select the α with the highest R^2 for P/X to determine the simplified model. For the case study an α of 0.02 is selected (Supplementary Figure S2).

S1.3 Validation Simulations

In addition to the experimental validation, we add additional validation with simulated data.

S1.3.1 Simulate experiments

For the validation simulations, we generated 400 randomly reduced model equations (initial models). Although this meant that some parameters of the general model were randomly missing, certain parameters are always present as they comprise the simplest viable process model:

- γ_{\max}^o and K_m^o to describe the uptake solely dependent on the substrate concentration without inhibition,
- γ_{\min}^α and K_g^α to describe a linear maintenance function dependent on the uptake rate, and
- γ_{\max}^π and K_m^π to describe the production, dependent on the available substrate flux without inhibition.

In addition to these 6 necessary parameters, the model can include up to 15 additional terms, 4 inhibiting uptake, 4 inhibiting production, 4 increasing maintenance, and 3 adding temperature dependence. Each temperature dependence adds three additional model parameters while other terms add one model parameter (see Table S1). For each possible number of additional terms ($i = 0$ to 15), we randomly select 25 sets of parameters (choosing i terms from the 15 available terms) and generate random values for these parameters. To get a reasonable range for inhibition and activation, the parameter value distribution was defined such that its median halves the product or uptake rate and doubles the maintenance requirement (detailed ranges are listed in Supplementary Table S5).

Subsequently, for all these initial models values for X , P , G , and V are calculated for all runs and all measurement points according to the DoE, used for the training data, with the model equations (1-4). The initial biomass was set to 30 g L^{-1} , i.e., the mean of the center point experiments. Parameterized models that do not reach a product concentration of at least 3 g L^{-1} in any process of the DoE are discarded and replaced by new equations created in the same way.

Next, we added randomly sampled errors to the biomass and product concentrations, mimicking measurement data with biological variance, sampling, and measurement errors.

$$\hat{X} = X(1 + \epsilon_X), \quad \epsilon_X \sim \mathcal{N}(0, \sigma_X), \quad \hat{P} = P(1 + \epsilon_P) \quad \epsilon_P \sim \mathcal{N}(0, \sigma_P) \quad (\text{S3})$$

where $\mathcal{N}(0, \sigma)$ is the normal distribution and $\sigma_X \in \{0, 0.015, 0.03, 0.06\}$ and $\sigma_P \in \{0, 0.02, 0.04, 0.08, 0.16, 0.32\}$. 0.03 and 0.08 are the observed errors in the training data (see S1.3.3). Other values are added to explore the effect of errors on the OptFed results.

S1.3.2 Compare OptFed results with initial models

For all simulated experimental datasets we use OptFed to fit the model, including selecting relevant parameters. For each initial model, we calculate the optimal feed rate and temperature with OptFed and with RSM (exponential feeds are used to reduce the computational effort and make OptFed results comparable to RSM results). This gives us three sets of control variables (μ_f and T), and for each one, we calculate the final product-to-biomass yield using the initial model. The ratio between the optimum achieved by using the model prediction and the real optimum for the initial model is used as the performance indicator for the RSM and OptFed.

$$r_{\text{opt}}^{\text{OptFed}} = \frac{\mathcal{Y}_{\bar{X}}^P \left(\mu_{f,\text{opt}}^{\text{OptFed}}, T_{\text{opt}}^{\text{OptFed}} \right)}{\mathcal{Y}_{\bar{X}}^P \left(\mu_{f,\text{opt}}, T_{\text{opt}} \right)} \quad r_{\text{opt}}^{\text{RSM}} = \frac{\mathcal{Y}_{\bar{X}}^P \left(\mu_{f,\text{opt}}^{\text{RSM}}, T_{\text{opt}}^{\text{RSM}} \right)}{\mathcal{Y}_{\bar{X}}^P \left(\mu_{f,\text{opt}}, T_{\text{opt}} \right)} \quad (\text{S4})$$

A ratio close to 1 means we get very close to the optimum regarding the target variable P/X .

S1.3.3 Estimating Variance in the training data

We require sensible estimates of the expected errors to simulate meaningful experimental data. This is done using the four center point runs of the training data. These runs have identical uptake rates (feed rate is adapted to biomass concentration and volume) but different initial biomass concentrations and volumes. We compensate for this by calculating adjusted product and biomass concentrations for the process i (X_{adj}^i and P_{adj}^i , $i \in \mathcal{P}$) independent of the initial biomass concentration (X_0) and the initial volume (V_0). For $x \in \{X, P\}$ we calculate

$$x_{\text{adj}}^i(t) = \frac{x^i(t)V^i(t)}{X_0^i V_0^i} \quad (\text{S5a})$$

and based on this the relative deviation from the mean at each time-point

$$x_{\text{err}}^i(t) = \frac{x_{\text{adj}}(t) - \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} x_{\text{adj}}^i(t)}{\frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} x_{\text{adj}}^i(t)} \quad (\text{S5b})$$

for each process the mean error and the variance are calculated over all measurements points \mathcal{M}

$$\overline{x_{\text{err}}^i} = \frac{1}{|\mathcal{M}|} \sum_{t_j \in \mathcal{M}} x_{\text{err}}^i(t_j) \quad (\text{S5c})$$

$$s^2(x_{\text{err}}^i) = \frac{1}{|\mathcal{M}| - 1} \sum_{t_j \in \mathcal{M}} (x_{\text{err}}^i(t_j))^2 \quad (\text{S5d})$$

and use the mean of the variance to characterize the random error expected within a process.

$$s^2(x_{\text{err}}) = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} s^2(X_{\text{err}}^i) \quad (\text{S5e})$$

The expected random error within one process in (S5e) is critical to fit the uptake, growth, and production rates and the model equations. We use this to assess the simulated validation results.

S1.4 Temperature stability

To estimate the stability of our temperature optimum we use a strategy based on the Monte Carlo method [64–66], using the estimated relative errors of our data determined in Section S1.3.3 (3% for biomass and 8% for product). We create 1000 disturbed datasets according to Eq. (S3) and fit each set's model for γ^π . For each model, we calculate the optimal temperature and intervals where p % of the maximum is reached ($0 < p < 100$). We now search for the maximal ratio and the corresponding temperature we can guarantee statistically significant ($\alpha = 0.05$). To do this, we determine the temperature for each ratio p where the least number of models is outside the temperature ratio. We search for the highest ratio ($\gamma^\pi(T)/\gamma^\pi(t_{\text{opt}})$) where more than 95% of all processes reach this ratio at the optimal temperature. This temperature is the stable optimum.

S2 Supplementary Tables

Supplementary Table S1: List of parameters used throughout the manuscript. Values shows the parameter values identified by OptFed. Where no value is given, this parameter was removed during model simplification. Increased error gives the additional fitting error (calculated as RSS, see (12)) if this parameter would be removed as well.

Name	Unit	Value	increased error	
Parameters (fitted using training data)				
<i>Parameters of base model</i>				
$c_{\gamma_{\max}^{\circ}}$	maximal uptake rate	$\text{g g}^{-1} \text{h}^{-1}$	0.49	
K_m°	dependence on substrate concentration	g	1.0×10^{-3}	
$c_{\gamma_{\min}^{\alpha}}$	maintenance requirement without growth	$\text{g g}^{-1} \text{h}^{-1}$	2.4×10^{-5}	
K_g^{α}	growth dependent maintenance	$\text{g g}^{-1} \text{h}^{-1}$	1.0×10^{-4}	142 %
$c_{\gamma_{\max}^{\pi}}$	maximal production rate	$\text{g g}^{-1} \text{h}^{-1}$	-	
K_m^{π}	production dependence on available substrate	$\text{g g}^{-1} \text{h}^{-1}$	0.62	188 %
inhibition parameter for γ°				
K_G°	substrate inhibition	g L^{-1}	89	119 %
K_n°	inhibition through no. of generations		-	
K_P°	product inhibition	g L^{-1}	-	
K_X°	biomass inhibition	g g^{-1}	-	
parameters for increased γ^{α}				
K_G^{α}	increase caused by free substrate	g L^{-1}	-	
K_n^{α}	increase with increasing no. of generations		-	
K_P^{α}	increase caused by product	g L^{-1}	0.11	37 %
K_X^{α}	increase caused by biomass	g g^{-1}	-	
inhibition parameters for γ^{π}				
K_G^{π}	effect of available substrate on production	g L^{-1}	-	
K_n^{π}	inhibition for higher no. of generations		1.5	50 %
K_P^{π}	Product inhibition	g g^{-1}	-	
K_X^{π}	Biomass inhibition	g L^{-1}	-	
<i>Temperature dependence for γ°</i>				
$\Delta G_{\text{cat}}^{\circ}$	catalytic activation energy	J mol^{-1}	-	
$\Delta H_{\text{eq}}^{\circ}$	enthalpic (conversion of active to inactive enzyme)	J mol^{-1}	-	
T_{eq}°	temperature where half the enzyme is active	K	-	
E_0°	hypothetical enzyme concentration	K	-	
<i>Temperature dependence for γ^{α}</i>				
$\Delta G_{\text{cat}}^{\alpha}$	catalytic activation energy	J mol^{-1}	-	
$\Delta H_{\text{eq}}^{\alpha}$	enthalpic (conversion of active to inactive enzyme)	J mol^{-1}	-	
T_{eq}^{α}	temperature where half the enzyme is active	K	-	
E_0^{α}	hypothetical enzyme concentration	K	-	
<i>Temperature dependence for γ^{π}</i>				
$\Delta G_{\text{cat}}^{\pi}$	catalytic activation energy	J mol^{-1}	5.2×10^4	
$\Delta H_{\text{eq}}^{\pi}$	enthalpic (conversion of active to inactive enzyme)	J 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}	
} 38 %				
<i>Constants (physical constants and yield determined by metabolomic model)</i>				
$Y_{X/G}^{\pi}$	biomass yield per substrate	g g^{-1}	0.627	
$Y_{P/G}^{\pi}$	product yield per substrate	g g^{-1}	0.652	
k_B	Boltzmann constant	J K^{-1}	1.38×10^{-23}	
R	Gas constant	$\text{J mol}^{-1} \text{K}^{-1}$	8.314	
h	Plank constant	J h	2.39×10^{-30}	

Supplementary Table S2: Maximal and minimal values for all fitted parameters.

Parameter	min	max
K_m°	1×10^{-15}	1×10^2
K_X°	0	1×10^4
K_P°	0	1×10^4
K_n°	0	1×10^4
K_G°	1×10^{-5}	1×10^4
$c_{\gamma_{\max}}^{\circ}$	0	1
E_0°	0	1×10^{-5}
$\Delta G_{\text{cat}}^{\circ}$	1×10^4	1×10^5
$\Delta H_{\text{eq}}^{\circ}$	1×10^5	1×10^7
T_{eq}°	300	315
K_g^{α}	1×10^{-6}	∞
K_X^{α}	1×10^{-6}	∞
K_P^{α}	1×10^{-6}	∞
K_n^{α}	1×10^{-6}	∞
K_G^{α}	1×10^{-6}	∞
$c_{\gamma_{\min}}^{\alpha}$	0	1
E_0^{α}	0	1×10^{-5}
$\Delta G_{\text{cat}}^{\alpha}$	1×10^4	1×10^5
$\Delta H_{\text{eq}}^{\alpha}$	1×10^5	1×10^7
T_{eq}^{α}	300	315
K_m^{π}	0	1×10^2
K_X^{π}	0	1×10^4
K_P^{π}	0	1×10^4
K_n^{π}	0	1×10^4
K_G^{π}	0	1×10^4
$c_{\gamma_{\max}}^{\pi}$	0	1
E_0^{π}	0	1×10^{-5}
$\Delta G_{\text{cat}}^{\pi}$	1×10^4	1×10^5
$\Delta H_{\text{eq}}^{\pi}$	1×10^5	1×10^7
T_{eq}^{π}	300	315

Supplementary Table S3: Estimated parameters for the response surface methodology (RSM) model according to Eqn. (11).

Parameter	c	c_f	c_T	c_{fT}	c_{f2}	c_{T2}
Value	-2.11×10^2	-16.7	1.39	0.0607	-7.45	-2.30×10^{-3}
p-Value	0.20	0.54	0.20	0.50	0.055	0.20

Supplementary Table S4: Performance summary of all processes (identified by ID) designed and analyzed in this work. For 3 temperatures, 31.0 °C (center point of training data), 33.5 °C, and 35.8 °C (optimal temperature) the optimum was calculated with linear feed function (opt. process) and free feed function.

ID	comment	feed rate, f [L h ⁻¹]	P [g L ⁻¹]	ΔP [%]	P/X [g g ⁻¹]	$\Delta P/X$ [%]	X_{end} [g L ⁻¹]	V/V_0
0	reference	-	7.76	0.00	0.14	0.00	57.11	1.42
1	measurements	-	7.67	-0.01	0.13	-0.04	58.47	1.29
— $T = 31.0$ °C —								
2	opt process	$0.02208 + 0.00377t$	7.51	-3.24	0.14	0.34	55.07	1.35
5	free feed	-	7.51	-3.23	0.14	0.34	55.07	1.35
— $T = 33.5$ °C —								
3	opt process	$0.02190 + 0.00443t$	8.65	11.47	0.16	18.99	53.50	1.38
6	free feed	-	8.65	11.48	0.16	18.99	53.50	1.38
— $T = 35.8$ °C —								
4	opt process	$0.02154 + 0.00526t$	9.64	24.23	0.19	37.08	51.76	1.42
7	free feed	-	9.64	24.22	0.19	37.08	51.75	1.42

Supplementary Table S5: Distribution of random parameter values for simulated models. The range of basic parameter values is based on the values observed in the training data. Typically an active parameter should halve the uptake or production rate for the highest observed values or double the maintenance. The temperature range is selected to reflect similar behavior to the reference processes with an optimum within the design space of the DoE.

Basic parameters:		
$c_{\gamma_{\max}^{\circ}}$	uniformly distributed	$[0.3, 1] \text{ g g}^{-1} \text{ h}^{-1}$
K_m°	uniformly distributed	$[0, 1] \text{ g}$
$c_{\gamma_{\min}^{\alpha}}$	uniformly distributed	$[0, 0.05] \text{ g g}^{-1} \text{ h}^{-1}$
K_g^{α}	uniformly distributed	$[0, \frac{1}{2\gamma_{\min}^{\alpha}}] (\text{g g}^{-1} \text{ h}^{-1})^{-1}$ (0 to 50 % of total uptake)
$c_{\gamma_{\max}^{\pi}}$	uniformly distributed	$[0, 0.1] \text{ g g}^{-1} \text{ h}^{-1}$
K_g^{α}	exponential distributed	mean = $1 \text{ g g}^{-1} \text{ h}^{-1}$
Temperature (for $i \in \{\mu, \pi, \circ\}$ $m \in \{\max, \min\}$):		
ΔG_{cat}^i	uniformly distributed	$[4 \times 10^4, 1 \times 10^5] \text{ mol J}^{-1}$
ΔH_{eq}^i	uniformly distributed	$[7 \times 10^4, 1 \times 10^6] \text{ mol J}^{-1}$
T_{eq}^i	uniformly distributed	$[304.15, 308.15] \text{ K}$
E_0^i	calculate from $c_{\gamma_j^i}$	$c_{\gamma_j^i} = \gamma_j^i (304.15) \text{ K}$
inhibition parameters for g		
K_G°	exponentially distributed	median = $\max(G) [\text{g L}^{-1}]$
K_n°	exponentially distributed	median = $\max(n)$
K_P°	exponentially distributed	median = $\max(P) [\text{g g}^{-1}]$
K_X°	exponentially distributed	median = $\max(X) [\text{g L}^{-1}]$
parameters for increased γ^{α}		
K_G^{α}	exponentially distributed	median = $(\max(G))^{-1} [\text{g L}^{-1}]$
K_n^{α}	exponentially distributed	median = $\max(n)$
K_P^{α}	exponentially distributed	median = $\max(P) [\text{g g}^{-1}]$
K_X^{α}	exponentially distributed	median = $\max(X) [\text{g L}^{-1}]$
inhibition parameters for γ^{π}		
K_G^{π}	exponentially distributed	median = $\max(G) [\text{g L}^{-1}]$
K_n^{π}	exponentially distributed	median = $\max(n)$
K_P^{π}	exponentially distributed	median = $\max(P) [\text{g g}^{-1}]$
K_X^{π}	exponentially distributed	median = $\max(X) [\text{g L}^{-1}]$

S3 Supplementary Figures

Author details

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

References

1. Stefan Kittler, Julian Ebner, Mihail Besleaga, Johan Larsbrink, Barbara Darnhofer, Ruth Birner-Gruenberger, Silvia Schobesberger, Christopher K Akhgar, Andreas Schwaighofer, Bernhard Lendl, et al. Recombinant protein I: production, purification and characterization of a universal binding ligand. *Journal of Biotechnology*, 359: 108–115, 2022.



















