Multi-fidelity Bayesian Optimisation of Syngas Fermentation Simulators

Mahdi Eskandari
Division of Computing, Engineering and Mathematical Sciences
University of Kent
Canterbury, United Kingdom
me377@kent.ac.uk

Lars Puiman
Faculty of Applied Sciences, Department of Biotechnology
Delft University of Technology
Delft, The Netherlands
L.Puiman@tudelft.nl

Jakob Zeitler
Matterhorn Studio
Oxford, United Kingdom
jakob@matterhorn.studio

Abstract

A Bayesian optimization approach for maximizing the gas conversion rate in an industrial-scale bioreactor for syngas fermentation is presented. We have access to a high-fidelity, computational fluid dynamic (CFD) reactor model and a low-fidelity ideal-mixing-based reactor model. The goal is to maximize the gas conversion rate, with respect to the input variables (e.g., pressure, biomass concentration, gas flow rate). Due to the high cost of the CFD reactor model, a multi-fidelity Bayesian optimization algorithm is adopted to solve the optimization problem using both high and low fidelities. We first describe the problem in the context of syngas fermentation followed by our approach to solving simulator optimization using multiple fidelities. We discuss concerns regarding significant differences in fidelity cost and their impact on fidelity sampling and conclude with a discussion on the integration of real-world fermentation data.

1 Introduction

Syngas fermentation is a recently emerged technology, with a promise as a sustainable and circular process for the production of the fuels and materials from a wide range of feedstocks, such as biomass, municipal solid waste streams, and atmospheric CO_2. LanzaTech, a prominent company in the field, has achieved successful commercialization of the fermentation process, wherein micro-organisms (*Clostridium autoethanogenum*) convert synthesis gas (comprising CO, H_2, and CO_2) from industrial off-gases into ethanol [Fackler et al., 2021, Köpke and Simpson, 2020]. While gas-to-liquid mass transfer limitations have historically posed challenges for scaling-up syngas fermentation, it is worth noting that such limitations can be significantly mitigated by producing bubble coalescence-suppressing products, such as ethanol [Puiman et al., 2022]. Industrial-scale (syngas) fermentation process modeling has been established as a pivotal tool for reactor scale-up Delvigne et al. [2017]. The models used for such applications range from relatively simple, and fast, 0D and 1D models with

a black-box description of the micro-organism to more complex 3D computational fluid dynamic (CFD) models (see e.g. Benalcázar et al. [2020], Siebler et al. [2019], Puiman et al. [2023]). A detailed description of the microbial behavior in the reactor could be obtained by coupling reactor models with metabolic models (e.g. Chen et al. [2018], Haringa et al. [2018]). Although CFD models provide a high-resolution description of the bioreactor, solving CFD models takes considerably more time (order of days to weeks) than solving 0D and 1D models (seconds to minutes). This is a major limitation for assessing the impact and the optimization of operating conditions, such as biomass concentration, pressure or gas composition, as such an iterative solving procedure in a CFD model could take months. Here, we present an approach for optimization of operating conditions using a Bayesian optimization (BO) framework [Frazier, 2018], and a 0D model, which was tested before on CFD modeling results [Puiman et al., 2023]. Specifically, we are using multi-fidelity BO, which allows us to flexible learn from both low and high-fidelity simulations [Poloczek et al., 2017].

2   Problem

Puiman et al. [2023] present two bioreactor models for industrial-scale syngas fermentation:

- A *high-fidelity*, CFD model, that takes a few (~ 5) days per run.
- A *low-fidelity*, ideal-mixing model that takes less than 2 minutes per run.

We are interested in finding the optimal operating parameters for industrial-scale syngas fermentation in these simulators, by efficiently exploring these two fidelities. The CFD model can be used to estimate the spatial variation in dissolved gas concentration and microbial uptake rates. It does so by solving the multiphase flow behavior in an Eulerian fashion, while the turbulence is solved for using a RANS (Reynolds-Averaged Navier-Stokes) model. Transport of CO, H_2, CO_2 is solved in both gas and liquid phases, while mass transfer and reaction are solved using the same methods as in the ideal-mixing model, but with high spatial distribution [Puiman et al., 2022, 2023]. In the CFD model, all these equations are solved for $370,000$ mesh elements during a period of 1000 s. One simulation usually lasts a week with 48 AMD EPYC cores. The ideal-mixing model can be used for fast calculations of the industrial-scale syngas fermentation bioreactor. It is a good initial approximation (as observed in Figure 3 from Puiman et al. [2023]) for the CFD model. Gas-to-liquid mass transfer and microbial syngas conversion are described by:

$$(k_L a)_i (H_i p y_i - c_{L,i}) = q_i c_X, \tag{1}$$

where

$$(k_L a)_i = k_{L,i} \frac{6 \epsilon_G}{d_B}, \quad \text{for } i = \text{CO}, \text{H}_2, \text{CO}_2, \tag{2}$$

$$\epsilon_G = \frac{u_{G,s}}{0.25}, \tag{3}$$

$$u_{G,s} = \frac{n_{G,s} R T}{A p}, \tag{4}$$

with the constants temperature $T = 273.15 K$, gas constant $R = 8.314 \frac{J}{\text{mol} K}$, reactor area $A = 19.63 m^2$, $k_{L,\text{CO}} = 3.98 \times 10^{-4} m/s$, $k_{L,\text{H}_2} = 5.93 \times 10^{-4} m/s$, $k_{L,\text{CO}_2} = 3.87 \times 10^{-4} m/s$ [Puiman et al., 2023]. The typical parameters used for the ideal-mixing model are listed in Table 1.

The following parameters were considered as a variable during operation:

1. Biomass concentration $c_X \in [0, 50] g/L$
2. Pressure $p \in [101325 - 506625] \text{Pa}$
3. Mole fraction y_i where $i \in \{CO_2, H_2, CO\}$. Note that the constraint $\sum_i y_i = 1$ applies for the mole fractions.
4. Bubble diameter $d_B \in [0, 10^{-3}] m$.
5. Flow in gas inlet $n_{G,s} \in [0, 200] \text{mol/s}$.

With the given parameters, the output of the model is the CO and H_2 reaction rate r_i, $i \in \{\text{CO}, \text{H}_2\}$ with $r_i = q_i c_X$. So the target is to calculate q_i, considering CO and H_2 uptake kinetics [Puiman

a) Industrial-scale syngas fermentation

b) High fidelity model
CFD model predicts high spatial variation of CO conversion rate

c) Low fidelity model
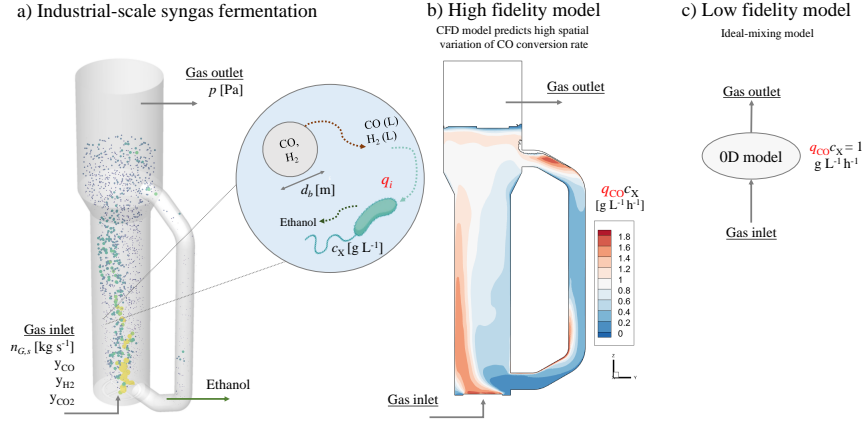Ideal-mixing model

Figure 1: Schematic descriptions of the used models. a) Conceptual representation of the industrial-scale reactor, and the operating parameters, b) result of a run of the high fidelity CFD model, c) result of the low-fidelity ideal-mixing model. The variable in red q_i is the one solved for in both modeling routines.

et al., 2023] and Equation 1.

$$q_{\mathrm{CO}} = q_{\mathrm{CO}}^{\max} \left(\frac{c_{\mathrm{L,CO}}}{K_{\mathrm{S,CO}} + c_{\mathrm{L,CO}} + \frac{c_{\mathrm{L,CO}}^2}{K_I}} \right) \tag{5}$$

$$q_{\mathrm{H_2}} = q_{\mathrm{H_2}}^{\max} \left(\frac{c_{\mathrm{L,H_2}}}{K_{\mathrm{S,H_2}} + c_{\mathrm{L,H_2}}} \right) \left(\frac{1}{1 + \frac{c_{\mathrm{L,CO}}}{K_{\mathrm{I,CO}}}} \right). \tag{6}$$

The parameters for solving (5) and (6) are listed in Table 1 in [Puiman et al., 2023]. The CFD (high fidelity) and ideal-mixing (low fidelity) model both take the same values as input and give $q_i \ i \in \{\mathrm{H_2, CO}\}$ as output. The comparison of the results of both simulators, for varying biomass concentrations c_X is provided in Figure 3 in [Puiman et al., 2023]. The difference between them as mentioned is that the high fidelity model takes more than two days to complete a run whereas the scale-down simulator takes less than a minute to provide results since its only complexity comes from solving a system of equations in (5) and (6). The ideal-mixing model, however, is a good estimation for the average q_{CO} and q_{H2}, while the CFD model also provides information on its spatial distribution inside the bioreactor. The goal is to maximize the syngas conversion rate r_i with respect to the input values. Since the industrial-scale simulator has a high complexity and cost for each experiment, directly maximizing its objective function is not feasible.

3 Solution

In order to solve the above optimization problem, we apply multi-fidelity Bayesian optimization (MFBO, [Poloczek et al., 2017]). Classic Bayesian optimization (BO, [Frazier, 2018]) is a machine learning technique for efficiently optimizing complex functions, particularly when evaluations are computationally expensive and there is no access to a gradient. It combines Bayesian statistical modeling with information-gain functions to iteratively refine its modeling of the objective function to guide the search for minima or maxima. Usually, BO is started with a small set of evaluations to build a *probabilistic model* (often a Gaussian Process, [Rasmussen and Williams, 2006]) that captures the function's behavior and uncertainty around model predictions. It then selects the next point to evaluate based on *acquisition functions*, which, depending on operator preference, explore uncertain regions or exploit promising areas. This optimisation loop continues iteratively, gradually converging towards the optimal solution(s) while trying to keep small the number of function evaluations required. BO is widely used in machine learning hyperparameter tuning, experimental design, and other domains where optimizing costly functions is a challenge. Since there are two fidelities available, i.e., high fidelity (the CFD model), and low fidelity (the ideal-mixing model), discrete MFBO [Poloczek
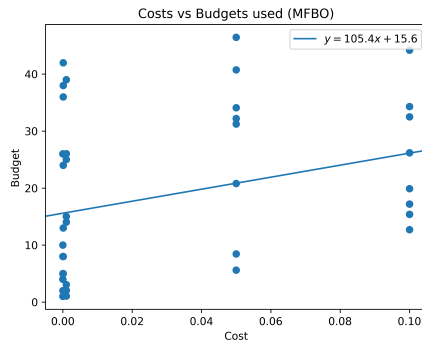
3

Figure 2: Linear regression of budget spend onto the cost of the single fidelity. The linear relationship is shown top right. The slope is statistically significant at a p-value of 0.038.

et al., 2017] is utilized to solve the problem. The goal of [Poloczek et al., 2017] is to maximize an expensive-to-evaluate function where a cheaper approximation of the black box is also available. The high fidelity model is modeled as $\mathcal{IS}_0$ whereas the low fidelity one is modeled as $\mathcal{IS}_1$. Each of these information sources is also associated with a cost function c_j with $j = \{0, 1\}$. It is assumed that the cost function is known and continuous. For instance, in the syngas fermentation model, the cost function can be modeled as the time required for each run. The optimization algorithm proceeds in rounds such that using the previously sampled data a posterior Gaussian process is fit to the initial data. Then, in each run an information source is picked that maximizes the cost-sensitive Knowledge Gradient in the equation in (7). Then with the new observation, the posterior is updated. The details of the misoKG algorithm are listed in [Poloczek et al., 2017] and are omitted for brevity.

$$\mathrm{MKG}^n(l, x) = \mathbb{E}_n \left[\frac{\max_{x' \in \mathcal{D}} \mu^{(n+1)}(0, x') - \max_{x' \in \mathcal{D}} \mu^{(n)}(0, x')}{c_l(x)} \Big| l^{(n+1)} = l, x^{(n+1)} = x \right]. \tag{7}$$

4 Discussion

Before we commit to an optimization campaign of unknown length, we wanted to understand the MFBO behavior better to adjust our expectations on how many high-fidelity simulator runs (each 2-5 days) we should expect. Eq. (7) shows the equation of the acquisition function used in our MFBO model. Most simply, it is a nested optimization problem, where we maximize the gain in the objective function for each information source $\mathcal{IS}_l$ and then choose the source with the highest gain overall. The MKG for each information source is weighted by the cost of acquiring a sample from it.

We have significant concerns regarding a situation in which low-fidelity queries would dominate, resulting in fewer requests for high-fidelity runs. This could lead to a cubic increase in the demand for modeling resources, posing a significant computational obstacle. A brief examination of Eq. 7 from [Poloczek et al., 2017] reveals that when the cost $c_l(x)$ decreases, the MKG value increases. This suggests that the model is more inclined to recommend that particular point as the next sample query. In fact, from an intuitive standpoint, if a low-fidelity information source is significantly less expensive than its high-fidelity counterpart while offering comparable performance, it is logical to predominantly depend on the low-fidelity source for experimentation.

Informed by the equation in Eq. 7, we did a brief empirical study to understand the influence of fidelity cost on information source selection. Presumably, the lower cost of the low fidelity will cause the acquisition function to suggest more queries of the low fidelity. We ran the standard BoTorch discrete MFBO tutorial [et al, 2021] on optimizing a multi-fidelity version of the Hartmann6D test function with costs of the lower fidelity set at $c_l(x) \in \{0.001, 0.05, 0.1\}$, 8 times each for different seeds, for 50 acquisition steps. We are interested to see how different low-fidelity costs impact the budget spend, both low and high fidelity added together, in the first 50 acquisition steps. Based on those samples, we then ran a Pearson correlation test and linear regression of the budget used for each trial and the cost $c_l(x)$. The hypothesis is that as the cost of the low fidelity increases, so does that budget-use over 50 iterations. We identified a positive correlation with a p-value of 0.038 which

generally is accepted as statistically significant and as such is in agreement with both theory and intuition laid out above, see Figure 2. Equipped with the intuition on how MFBO will query low and high-fidelity via the above empirical study, we are now confident as a next step to engage in an MFBO campaign to optimize the high-fidelity simulator with the help of low-fidelity samples. A possible extension would be a modification of the acquisition function would tie the execution time of the two fidelities into the cost and down-weight the querying of low-fidelities.

Ultimately, we hope to incorporate real-world fermentors as supplementary high-fidelity information sources alongside the existing high-fidelity simulator. Fermentation engineers can utilize the publicly available MFBO code [1] to optimize their fermentation conditions. They can pool both low and high-fidelity simulators such as in [Puiman et al., 2023], as well as their large-scale fermentation data, to maximize the reaction rate of their syngas bioreactors.

5   Conclusion

We presented a multi-fidelity Bayesian optimization approach for the optimization of syngas conversion in industrial-scale bioreactors. With real-world experimentation often prohibitively expensive, simulation approaches are an indispensable choice for bioreactor engineers tasked with scale-up. Pooling the data between simulators of different fidelity and costs can make a significant difference to fermentation engineering. We anticipate that our computational approach presented here can guide the optimization search in both academia and industry. We hope that it will help identify the optimal operating conditions for bioreactors as part of the critical contribution that syngas fermentation can make toward a sustainable planet.

6   Acknowledgments

Appendix: List of symbols

Table 1: List of the parameters used for simulating the ideal-mixing model.

Symbol	Description	Unit
A_p	Reactor area	m^2
c	Concentration in liquid phase	$\mathrm{g\,L^{-1}}$ or $\mathrm{mol\,m^{-3}}$
d_B	Bubble diameter	m
ϵ_G	Gas hold-up	-
H	Henry constant	$\mathrm{mol\,m^{-3}\,Pa^{-1}}$
K_i	Inhibition constant	-
k_L	Mass transfer coefficient	$\mathrm{m\,s^{-1}}$
$k_L a$	Volumetric mass transfer coefficient	$\mathrm{s^{-1}}$
K_s	Half-saturation constant	-
$n_{G,s}$	Molar gas flow	-
p	Pressure	Pa
q	Biomass-specific rate	$\mathrm{mol}_i\,\mathrm{mol}_X^{-1}\,\mathrm{s}^{-1}$
R	Gas constant	$\mathrm{J\,mol\,K}$
T	Temperature	K
$u_{G,s}$	Superficial gas velocity	$\mathrm{m\,s^{-1}}$
y	Gas fraction	$\mathrm{mol}_i\,\mathrm{mol}_G^{-1}$
i	Compound	-
L	Liquid phase	-
X	Biomass	-

---

[1] See tutorial and single-click pipeline access here: `https://matterhorn.studio/pages/seminars/syngas-fermentation-optimisation/`

References

Nick Fackler, Bjorn D. Heijstra, Blake J. Rasor, Hunter Brown, Jacob Martin, Zhuofu Ni, Kevin M. Shebek, Rick R. Rosin, Seacutean D. Simpson, Keith E. Tyo, Richard J. Giannone, Robert L. Hettich, Timothy J. Tschaplinski, Ching Leang, Steven D. Brown, Michael C. Jewett, and Michael Köpke. Stepping on the Gas to a Circular Economy: Accelerating Development of Carbon-Negative Chemical Production from Gas Fermentation. *Annual Review of Chemical and Biomolecular Engineering*, 12:439–470, jun 2021. doi: 10.1146/ANNUREV-CHEMBIOENG-120120-021122.

Michael Köpke and Séan D. Simpson. Pollution to products: recycling of 'above ground' carbon by gas fermentation. *Current Opinion in Biotechnology*, 65:180–189, oct 2020. doi: 10.1016/j.copbio.2020.02.017.

Lars Puiman, Britt Abrahamson, Rob G.J.M. van der Lans, Cees Haringa, Henk J. Noorman, and Cristian Picioreanu. Alleviating mass transfer limitations in industrial external-loop syngas-to-ethanol fermentation. *Chemical Engineering Science*, 259:117770, sep 2022. doi: 10.1016/J.CES.2022.117770.

Frank Delvigne, Ralf Takors, Rob Mudde, Walter van Gulik, and Henk Noorman. Bioprocess scale-up/down as integrative enabling technology: from fluid mechanics to systems biology and beyond. *Microbial Biotechnology*, 10(5):1267–1274, sep 2017. ISSN 1751-7915. doi: 10.1111/1751-7915.12803.

Almeida Eduardo Benalcázar, Henk Noorman, Rubens Maciel Filho, and John A Posada. Modeling ethanol production through gas fermentation: a biothermodynamics and mass transfer-based hybrid model for microbial growth in a large-scale bubble column bioreactor. *Biotechnology for Biofuels*, 13(59):1–19, 2020. doi: 10.1186/s13068-020-01695-y.

Flora Siebler, Alexey Lapin, Maria Hermann, and Ralf Takors. The impact of CO gradients on C. ljungdahlii in a 125 m3 bubble column: Mass transfer, circulation time and lifeline analysis. *Chemical Engineering Science*, 207:410–423, nov 2019. doi: 10.1016/j.ces.2019.06.018.

Lars Puiman, Eduardo Almeida Benalcázar, Cristian Picioreanu, Henk J Noorman, and Cees Haringa. Downscaling industrial-scale syngas fermentation to simulate frequent and irregular dissolved gas concentration shocks. *Bioengineering*, 10(5):518, 2023. doi: 10.3390/bioengineering10050518.

Jin Chen, James Daniell, Derek Griffin, Xueliang Li, and Michael A. Henson. Experimental testing of a spatiotemporal metabolic model for carbon monoxide fermentation with Clostridium autoethanogenum. *Biochemical Engineering Journal*, 129:64–73, jan 2018. doi: 10.1016/j.bej.2017.10.018.

Cees Haringa, Wenjun Tang, Guan Wang, Amit T. Deshmukh, Wouter A. van Winden, Ju Chu, Walter M. van Gulik, Joseph J. Heijnen, Robert F. Mudde, and Henk J. Noorman. Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization. *Chemical Engineering Science*, 175:12–24, jan 2018. doi: 10.1016/j.ces.2017.09.020.

Peter I. Frazier. A tutorial on bayesian optimization, 2018.

Matthias Poloczek, Jialei Wang, and Peter Frazier. Multi-information source optimization. *Advances in neural information processing systems*, 30, 2017.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning.* Adaptive computation and machine learning. MIT Press, 2006. ISBN 026218253X.

Balandat et al. Multi-fidelity bo with discrete fidelities using kg. 2021. URL `https://botorch.org/tutorials/discrete_multi_fidelity_bo`.