11th International Conference on Sensitivity Analysis of Model Output, April 23-25 2025, Grenoble, France

Reduced-space Bayesian optimization of process flowsheets

NIKI TRIANTAFYLLOU[a,b], BEN LYONS[a,b], ANDREA BERNARDI[a,b], BENOIT CHACHUAT[a,b], CLEO KONTORAVDI[a,b], MARIA M. PAPATHANASIOU[a,b]

[a]*The Sargent Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

[b]*Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

Mathematical models are essential for evaluating and optimizing the performance of complex processes, providing insights into key performance indicators (KPIs) such as cost, efficiency, and environmental impact. However, these models often function as black-box systems, where underlying equations and derivative information are unavailable, making optimization challenging. This is particularly true for commercially available simulators that rely on steady-state models and heuristic rules, which complicate the identification of optimal process configurations. In such cases, data-driven optimization techniques, like Bayesian Optimization (BO), become highly valuable.

Bayesian Optimization (BO) is well-suited for handling expensive, black-box models (Brochu et al., 2010). However, BO may be challenged by high-dimensional problems. In this context, Global Sensitivity Analysis (GSA) can help identify the most influential variables that drive variability in the objective function (Saltelli et al., 2010). By quantifying both individual and interaction effects, GSA can reduce the variable set to only the critical ones (Kucherenko, 2013), improving BO efficiency. The integration of BO with GSA still faces challenges in balancing computational efficiency and solution accuracy, particularly when applied to complex simulation-based models. Additionally, surrogate-based models, while faster to evaluate, may not fully capture the true objective function, leading to suboptimal solutions.

In this work, we compare simulation-based and surrogate-based Bayesian optimization (Triantafyllou et al., 2024). For the simulation-based methods, we first apply Bayesian Optimization (BO) to the full set of decision variables. Next, we incorporate GSA as a dimensionality reduction step, followed by BO on the reduced variable set. For the surrogate-based approaches, we begin by using GSA to identify the key variables that significantly influence the objective function. These variables are then used to train feed-forward neural networks (ANNs), resulting in simpler, lower-dimensional surrogate models. We then optimize the ANNs using two different approaches: BO and mixed integer programming (MIP) with a big-M reformulation of ReLU ANNs (Triantafyllou et al., 2024; Ceccon et al., 2022).

The performance of both simulation-based and surrogate-based methods is evaluated using two benchmark case studies with different flowsheet simulators: (a) plasmid DNA production in SuperPro Designer with 18 decision variables, and (b) dimethyl ether (DME) production in Aspen HYSYS with 14 decision variables. Generally, simulation-based methods yield superior solutions since they evaluate the true objective function at each step, whereas surrogate-based approaches optimize an approximation (ANN) of the true objective function. This can lead to the ANN having optima that differ from the true objective function both globally and locally (Figure 1). However, the Bayesian optimization of the ANN (GSA-ANN-BO) consistently demonstrates the fastest execution times, achieving reductions of two to three orders of magnitude compared to simulation-based BO, without accounting for the time taken for initial sampling. This makes it particularly advantageous for real-time and resource-constrained optimization tasks, where computational efficiency is critical without a significant loss in solution quality.

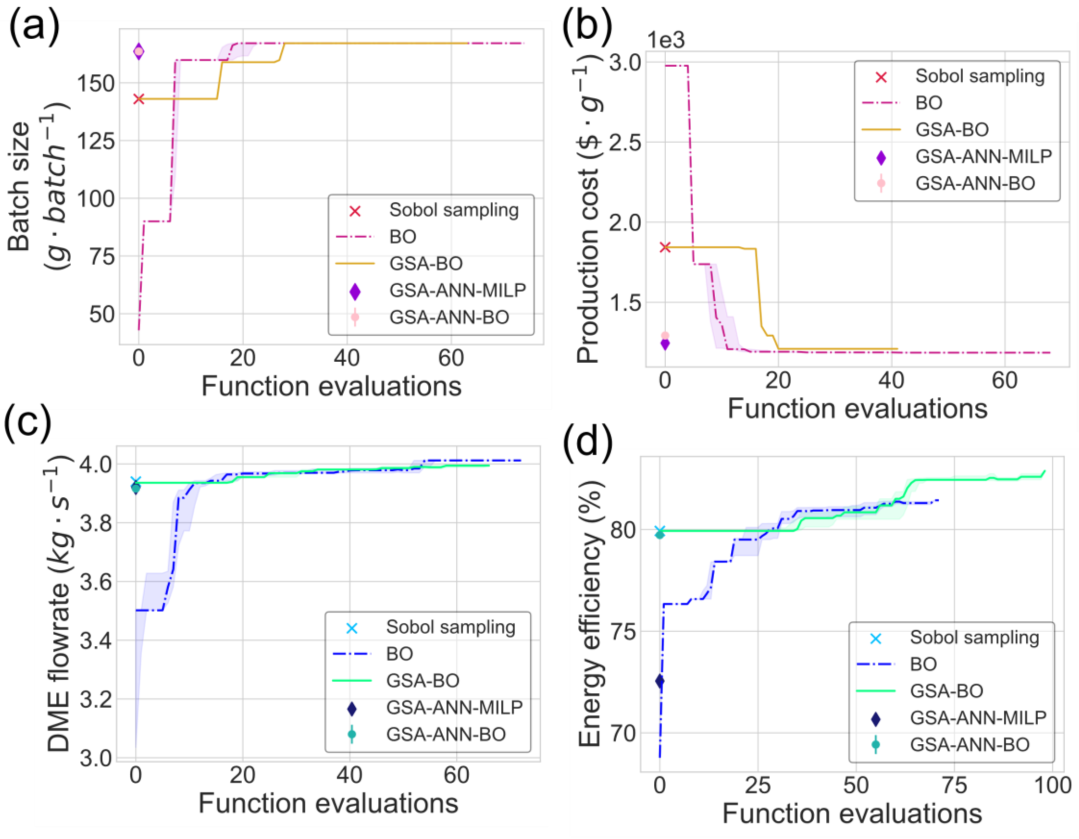

Figure 1. Overview of the optimization methods applied to manufacturing process optimization across the SuperPro Designer (a, b) and Aspen HYSYS (c, d) case studies (adapted from Triantafyllou et al., 2024). The simulation-based approaches include pure Bayesian optimization (BO) and Bayesian optimization combined with global sensitivity analysis (GSA-BO). The surrogate-based approaches consist of GSA-enhanced neural networks optimized using mixed-integer linear programming (GSA-ANN-MILP) and Bayesian optimization of GSA-enhanced neural networks (GSA-ANN-BO). For the Bayesian optimization methods, the plots display the median values along with confidence intervals (1st and 3rd quartiles) based on 10 random seed runs. The best value identified using Sobol sampling is also shown for comparison.

References :

[1] E. Brochu, V.M. Cora, N. De Freitas, "A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," arXiv preprint arXiv:1012.2599, 2010.

[2] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index," Comput. Phys. Commun., 181(2):259-270, 2010.

[3] S. Kucherenko, "SobolHDMR: a general-purpose modeling software," Synthetic Biology, 191-224, 2013.

[4] N. Triantafyllou, B. Lyons, A. Bernardi, B. Chachuat, C. Kontoravdi, M.M. Papathanasiou, "Comparative assessment of simulation-based and surrogate-based approaches to flowsheet optimization using dimensionality reduction," Comput. Chem. Eng., 189:108807, 2024.

[5] F. Ceccon, J. Jalving, J. Haddad, A. Thebelt, C. Tsay, C.D. Laird, R. Misener, "OMLT: Optimization & machine learning toolkit," J. Mach. Learn. Res., 23(349):1-8, 2022.