**Systematic comparison of flowsheet optimization options: surrogate modelling vs. genetic algorithms vs. Bayesian optimization**

Caroline Ganzera\*, Kai Sundmachera,b

aMax Planck Institute for Dynamics of Complex Technical Systems, Department of Process Systems Engineering, Sandtorstraße 1, 39106 Magdeburg, Germany

bOtto von Guericke University, Chair for Process Systems Engineering, Universitätsplatz 2, 39106 Magdeburg, Germany

\*cganzer@mpi-magdeburg.mpg.de

Abstract

This project aims to systematically compare options for flowsheet optimization. Surrogate modelling, genetic algorithms, Bayesian optimization, and combinations of them can be used to approach the black-box optimization task which arises from complex process simulation. There are trade-offs in computational effort, number of hyperparameters, reliability in reaching global optima, etc. The goal is to categorize existing methods, evaluate the performance of the algorithms, and synthesize guidelines on which algorithms to use depending on the flowsheet, objective, and dimensionality.

**Keywords**: surrogate modelling, Bayesian optimization, genetic algorithms, flowsheet optimization, derivative-free optimization, gradient-free optimization

* 1. Background

Process modelling with flowsheet simulators is a powerful, established tool for process design. When the formulation of the simulation is equation-based, optimization of the flowsheet can be solved with deterministic, gradient-based optimization algorithms. In many cases, however, gradients are not available from these simulations, calling for derivative-free optimization (DFO). Further, they frequently exhibit high computational cost, resulting in the need to limit the number of function evaluations. One established method to approach this task is surrogate modelling. A data set is generated from the simulation, which is then used to train a surrogate model, such as an artificial neural network (ANN). As the structure of the surrogate model is known, deterministic optimization is used to optimize over the surrogate model. Alternatively, genetic algorithms (GA) can be applied to flowsheets. It is a heuristic technique which uses evolving generations of testing points to iterate closer to optimal solutions. Recently, Bayesian optimization (BO) has gathered attention, and has been applied to flowsheet optimization (Sanchez Medina et al., 2021a). BO utilizes a surrogate model, often a Gaussian process (GP) model, and an acquisition function to estimate where sampling would reveal the most information. Many variations and even combinations of these methods exist (Sanchez Medina et al., 2021a, Sanchez Medina et al., 2021b). They vary in their accuracy, workflow complexity, computational effort, number of and sensitivity with respect to hyperparameters, ability to deal with constraints, and behavior at higher dimensions (degrees of freedom).

It is not obvious which type of DFO is appropriate or optimal depending on the flowsheet. Among the options, surrogate modelling has been a major focus (McBride and Sundmacher, 2019). The data requirement of effective ANN training can be a restrictive factor. GA tends to be applied much less frequently, but can be favored for multi-objective optimization (Shafiee et al., 2017). This can be helpful for chemical engineering applications, where there are often conflicting objectives such as efficiency, yield, cost, energy consumption, environmental impact. Its conception as heuristics-based may present a disadvantage. BO appears most helpful when function evaluations, i.e., simulation runs, are particularly expensive, but it comes at higher computational cost itself (van de Berg et al., 2022).

* 1. Methodology

We are evaluating the various options for flowsheet optimization, using ASPEN for flowsheet simulation. We plan to assemble a collection of DFO methods applicable to flowsheets as well as a set of different flowsheets, e.g., methanol synthesis from (green) H2 and CO2. Appropriate methods such as early stopping are utilized for training the surrogate models to ensure prediction accuracy and prevent overfitting. The set of flowsheets is designed deliberately diverse, with variation in types and number of unit operations as well as presence and number of recycles. Then the parameters characterizing the algorithms and the flowsheet optimization tasks can be varied and the performance of the algorithms can be estimated.

* 1. Conclusions

We expect trade-offs between the techniques in computational effort, number of hyperparameters, reliability in reaching global optima, etc. Further, the algorithms’ behavior at increasing number of degrees of freedom, and for varying objectives will be examined. Can some algorithms deliver adequate performance without the need to carefully adapt them to the flowsheet and tune the hyperparameters? What is the trade-off between adaptability and complexity? The goal is to categorize existing methods, evaluate the performance of the algorithms, and synthesize guidelines on which algorithms to use depending on the flowsheet, objective, dimensionality, and other factors.

References

E. I. Sanchez Medina, D. F. Rodriguez-Vallejo, E. A. del Rio-Chanona, K. Sundmacher, and P. Petsagkourakis, 2021a, Multi-objective Bayesian optimization of process flowsheets using trust regions and quality set metrics, 2021 AIChE Annual Meeting.

E. I. Sanchez Medina, D. F Rodriguez Vallejo, B. Chachuat, K. Sundmacher, P. Petsagkourakis, and E.A. del Rio-Chanona, 2021b, Acyclic modular flowsheet optimization using multiple trust regions and Gaussian process regression, Computer Aided Chemical Engineering, 50: 1117–1123.

K. McBride and K. Sundmacher, 2019, Overview of surrogate modeling in chemical process Engineering, Chemie Ingenieur Technik, 91(3):228–239.

A. Shafiee, M. Nomvar, Z. Liu, and A. Abbas, 2017, A new genetic algorithm based on prenatal genetic screening (PGS-GA) and its application in an automated process flowsheet synthesis problem for a membrane based carbon capture case-study, Chemical Engineering Research and Design, 128:265–289.

D. van de Berg, T. Savage, P. Petsagkourakis, D. Zhang, N. Shah, and E. A. del Rio-Chanona, 2022, Data-driven optimization for process systems engineering applications. Chemical Engineering Science, 248:117135.