Optimizing the Computational Time for Surrogate Model Training in Complex Chemical Process Simulations: An Application of the Traveling Salesman Problem

Luis Felipe Sánchez[a], Mattia Vallerio[a], Flavio Manenti[a]

[a] CMIC Department "Giulio Natta", Politecnico di Milano, Piazza Leonardo da Vinci 32, Milano, 20133

E-mail: mattia.vallerio@polimi.it

Process Simulations are key in the Chemical Engineering practice to emulate the behaviour of industrial-scale processes without requiring real-life experimentation. Often, the complexity of these simulations can be such that applications like process optimization or sensitivity analysis may prove unfeasible due to the long computational times required. To accelerate these calculations while maintaining accuracy, simpler data-driven models, like surrogate models, have been proposed in the literature. These models are trained by correlating input values, determined with a Design of Experiments (DoE) strategy, to the response or output of the simulation. Naturally, the objective is to extract the most information from the simulation with the smallest computational effort. With this objective in mind, the literature has focused on minimizing the number of required experiments with complex DoE strategies. However, the order in which these simulations are executed may also impact the total computational time. This is because for every set of input values, simulators solve complex systems of differential equations iteratively, resembling an error minimization problem. These problems are normally improved by providing an appropriate initial guess, which, in this context, may be a simulation with similar characteristics. Therefore, the problem can be reframed as finding an optimal sequence for executing the experiments to minimize the total Euclidean distance travelled in the input domain. This describes a widely studied optimization problem called the Travelling Salesman Problem (TSP), which has been addressed using several algorithms. In particular, the state-of-the-art ones in terms of computational efficiency and accuracy are the memetic (MA), the Lin-Kerninghan-Helsgaun (LKH), the Large Neighbour Search (LNS), and the Basic Iterated Local Search (BILS) algorithms. This work implements these algorithms to sort a series of Latin Hypercube (LHS) DoEs describing from one to ten input variables in a rigorous Aspen HYSYS simulation of a methanol production process from syngas. The study compares the computational time required to simulate each sequence of experiments to the original random sequence and provides insights regarding the applicability and efficiency of every TSP algorithm depending on the problem dimensionality. Future work will focus on expanding the number of algorithms and the problem dimensionality for the generalization of the methodology.

Keywords: *Surrogate Modeling, Optimization, Design of Experiments, Travelling Salesman Problem.*