Contribution ID: **32**   Type: **not specified**

Traditional or Adaptive Experimental Design? A Comparison of Statistical Design of Experiments and Bayesian Optimization for a Chemical Synthesis Problem

Friday, 30 May 2025 11:55 (20 minutes)

Bayesian Optimization (BO) has been recently shown as an efficient method for data-driven optimization of expensive and unknown functions. BO relies on a probabilistic surrogate model, commonly a Gaussian Process (GP), and an auxiliary acquisition function that balances exploration and exploitation for a goal-oriented experimental design, with the aim of finding the global optimum under a reduced experimental budget.

Before BO gained popularity, traditional statistical Design of Experiments (DOE) methods were commonly used to tackle such problems, such as, but not limited to, classic full/fractional factorial, response surface designs (central composite or Box-Behnken), and optimal designs. There are also other more recent developed methods such as Definitive Screening Designs or Orthogonal Minimally Aliased Response Surface (OMARS) designs. In contrast to BO, these designs are generally focused on the estimation of a quadratic response surface model and are mainly static in nature, although several rounds of experiments are typically recommended in practice where previous rounds' results are incorporated before the next round of experiments is determined. Nonetheless, there is a wide body of work, both theoretical and application driven, with success stories in different fields.

In this work, we consider a real-world chemical synthesis problem, where the goal is maximizing the reaction yield by manipulating a set of continuous and categorical variables, with a total of 22 variables, after One-Hot-Encoding (OHE). This complex case study provides an appropriate scenario for comparing traditional and adaptive experimental design approaches for optimization.

For the surrogate model in BO, we evaluate both using the OHE approach and the use of chemical descriptors as informative features of the categorical variables. Due to a high-dimensional design space, we also consider the use of Sparse Axis-Aligned Subspace Bayesian Optimization (SAASBO), a fully Bayesian approach which places strong sparsity priors on the GP model to avoid overfitting and relies on Markov Chain Monte Carlo for inference.

We then compare the BO approaches with a sequential statistical DOE approach, where an initial design is performed for variable screening, followed by a second and more efficient design on the variables deemed most important . The advantages and disadvantages of the different methods are highlighted in terms of implementation difficulty, rate of convergence to the optimum as well as model fit quality.

Type of presentation

Contributed Talk

Primary author: COUTINHO, João (Department of Chemical Engineering, University of Coimbra)

Co-authors: PENG, You (The Dow Chemical Company); RENDALL, Ricardo (Dow); RIZZO, Caterina (Eindhoven University of Technology/Dow); Dr MA, Kaiwen (The Dow Chemical Company); CHIN, Swee-Teng (The Dow Chemical Company); CASTILLO, Ivan (The Dow Chemical Company); P. SEABRA DOS REIS, Marco (Department of Chemical Engineering, University of Coimbra)

Presenter: COUTINHO, João (Department of Chemical Engineering, University of Coimbra)

Session Classification: Session

Track Classification: Spring Meeting