



Contribution ID: 129

Type: **not specified**

Quality by Design: Data-Efficient Active Learning Approaches

As referred in last year's abstract by the same authors, "the landscape of the pharmaceutical industry is evolving". Such a process continues, with more efforts being devoted to developing data-efficient methodologies for exploring operational spaces of increasing dimensionality that may be composed of continuous, categorical, and mixture factors. From what was (and still is) a science-based discipline, more awareness exists nowadays of the opportunities arising from exploring data-driven methodologies to conduct various key activities, namely to find the range of "optimal" conditions to operate a process or the best formulation for a pharmaceutical product. In this regard, two classes of active learning (AL) approaches are increasingly regarded as the most competitive: Statistical Design of Experiments (DOE) and Bayesian Optimisation (BO). Each one has emerged from a different scientific community –applied statistics and machine learning, respectively–, and has been conquering the confidence of supporters who, at the same time, build a diminished vision of the other class. DOE supporters tend to find BO lacking a sound theoretical structure and optimality guarantees, whereas BO proponents view the DOE approach as outdated and overly constrained by assumptions.

Ultimately, "the proof of the pudding is in the eating". Therefore, in this work, we present results on the use of both classes of methods in the sequential search for the best conditions, under a limited budget of experiments. Different systems are considered, and Monte Carlo simulations conducted to gather robust information on scenarios where one methodology is expected to outperform the other with high confidence. Drawing such an operational map of AL methods would, we argue, be a more useful outcome for practitioners than the entrenched defense of each class.

Acknowledgements

This work was funded by the CInTech project - Technological Hub for Innovation, Translation and Industrialisation of Complex Injectable Drugs -, under reference no. C644865576-00000005, co-financed by Componente C5 - Capitalização e Inovação Empresarial integrada na Dimensão Resiliência do Plano de Recuperação e Resiliência (PRR), through the NextGenerationEU fund. Authors also acknowledge support from CERES –Chemical Engineering and Renewable Resources for Sustainability Research Center, funded by FCT –Fundação para a Ciência e Tecnologia (UID/00102/2025), PRR –Recovery and Resilience Program, of the Portuguese Republic (UID/PRR/00102/2025), Equipar+2 (UID/PRR2/00102/2025).

Special/ Invited session

Classification

Both methodology and application

Keywords

Quality by Design; Sequential Design of Experiments; Bayesian Optimization; Active Learning

Primary author: P. SEABRA DOS REIS, Marco (Department of Chemical Engineering, University of Coimbra)

Co-author: VIDINHA BATISTA, Daniel Alexandre (University of Coimbra)

Presenter: P. SEABRA DOS REIS, Marco (Department of Chemical Engineering, University of Coimbra)

Track Classification: Design of Experiments