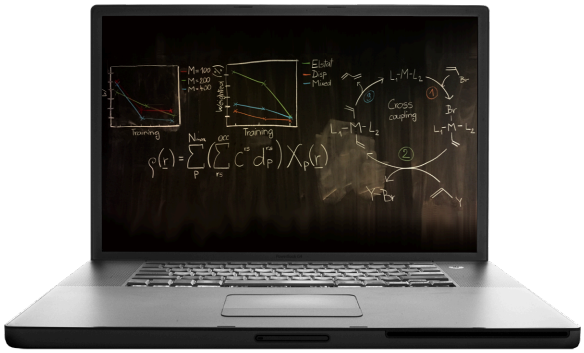
Data-driven Approaches to the Discovery and Classification of Homogeneous Catalysts

Clemence Corminboeuf, Matthew D. Wodrich, Boodsarin Sawatlon, Alberto Farbrizo

Laboratory for Computational Molecular Design, Institute of Chemical Sciences and Engineering, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland
clemence.corminboeuf@epfl.ch

With the goal of optimizing reactions, we exploit statistical learning techniques in tandem with quantum chemistry data to identify improved homogeneous catalysts. The appeal of machine-learning (ML) is quite clear; the computational speed up provided allows the properties and energetics of thousands of prospective catalysts to be rapidly assessed while preserving quantum chemical accuracy[1]. The massive quantity of data generated can be compiled and mined into an interactive tool, which facilitates the analysis and assists in identifying the most compatible metal/ligand family combinations with the goal of establishing relationships between the intrinsic chemical properties of different catalysts and their overall catalytic performance. We focus on both prototypical classes reactions and challenging organic processes and highlight some of the pro and cons of the overall approach associated with both the machine-learning strategies and the underlying screening strategies.



[1] Manuel Cordova, Matthew D. Wodrich, Benjamin Meyer, Boodsarin Sawatlon, Clemence Corminboeuf, *ACS Catalysis*, **2020**, *10*, 7021.
[2] Matthew D. Wodrich, Alberto Fabrizio, Benjamin Meyer, Clemence Corminboeuf *Chem. Sci.* **2020**, DOI: 10.1039/d0sc04289g.