TITLE: Enumerating the inorganic universe of small complexes for machine learning

AUTHORS (FIRST NAME, LAST NAME): Stefan O. Gugler[1,2], Jon Paul Janet[1], Heather J. Kulik[1]

INSTITUTIONS (ALL):
1. Chemical Engineering, Massachusetts Institute of Technology, Boston, MA, United States.
2. Department of Chemistry and Applied Biosciences, Swiss Federal Institute of Technology, Zürich, Switzerland.

ABSTRACT BODY:

Abstract: Transition metal complexes form promising functional inorganic materials due to their wide range of tunable electronic properties. However, exhaustive enumeration and calculation of all possible ligand fields is clearly intractable due to the vast nature of chemical space. Virtual high-throughput screening with density functional theory (DFT) allows us to harvest leads with desired properties but is severely constrained by 1) long calculation times and 2) variable accuracy. More accurate correlated methods are available to address 2) but drastically worsen 1). Machine learning techniques potentially allow us to address both issues simultaneously. Our group has previously developed data-driven models based on DFT results which have highlighted the dominant role of metal-proximal atoms (i.e. from the first and second coordination shell) in predicting spin state ordering, bond lengths, and ionization potential of the metal center. This motivates a systematic exploration of the space of octahedral complexes made of organic ligands with up to two heavy atoms (CNOPS), representing the metal-proximal environment. Even in this limited space, the number of potential candidate complexes is infeasible to calculate and so we propose a family of scoring functions that are used to extract mono- and bidentate ligands that most likely form stable complexes based on valency, net charge, and steric effects. The resulting organic ligand universe is then compared to similar studies of small organic molecules. Exploiting isoelectronic structure and empirical stability learned from previous studies, we sample the most promising compounds from this space with high-throughput DFT. We assess DFT performance selectively with more accurate correlated wavefunction calculations using domain-based local pair-natural orbital coupled cluster (DLPNO-CCSD(T)) and apply machine learning to model the difference between correlated wavefunction and DFT results in a composition-dependent manner. By doing this, we hope to learn property estimates for the full space of possible metal-proximal environments along with estimates of DFT reliability relative to DLPNO-CCSD(T).
(No Image Selected)