Testing Reproducibility in Materials Chemistry via Literature Meta-Analysis

David Sholl
John F. Brock III School Chair
School of Chemical & Biomolecular Engineering
Georgia Institute of Technology
The Need For Reliable Data in Chemical Engineering

Dow PDH plant in Freeport, TX
Approximate cost: $1 billion

- Far more energy efficient processes are possible, but perceived risk in an implemented process must be near zero
- Availability of physical data is a major impediment to design of new processes
- Little attention has been given to reproducibility in a formal sense

Walton and Sholl, *Joule* 2017
Common characteristics in applied materials chemistry problems
1. Many ($10^3$-$10^6$) possible materials
2. Well defined metrics used to judge which material is “best”
3. Potential IP and commercial value create complications for data sharing
Observation: In ‘hot’ topics, multiple groups study similar materials at similar times.

Implication: Replicate experiments may exist in literature even though replication was not the aim of the experiments.
What Is A Replicate (In Materials Chemistry)?

**Bronze**
Group A synthesizes material and measures properties twice

**Silver**
Group A synthesizes material
Groups A and B independently measure properties

**Gold**
Group A synthesizes material and measures properties
Group B independently synthesizes materials and measures properties
1000s of papers published reporting gas adsorption in MOFs

Great interest in CO$_2$ for applied and societal reasons

MOFs are crystalline materials, so materials are described in definitive terms

Measuring CO$_2$ uptake is “routine” using commercial instruments

Comprehensive database of experimental adsorption data has been compiled by NIST (adsorbents.nist.gov)

~13,000 isotherms recorded for ~300 molecules in ~6,000 different materials

Compiling a database of this type is laborious
Example 1: CO$_2$ Adsorption in HKUST-1 at Room Temperature

- CO$_2$ adsorption in HKUST-1 at 298 ± 5 K

Park, Howe and Sholl, *Chemistry of Materials* 2017 (DOI: 10.1021/acs.chemmater.7b04287)

18 replicates (27% classified as outliers)
Example 2: CO$_2$ Adsorption in UiO-66

Park, Howe and Sholl, *Chemistry of Materials* 2017 (DOI: 10.1021/acs.chemmater.7b04287)

Different representations of data depending on number of replicates

The methods we have introduced can be translated directly to other topics in materials chemistry and related fields.
What Do We Know About CO₂ Adsorption In MOFs?

- Only 9 materials have 4 or more replicates
- ~20% of CO₂ isotherms were classified as outliers
- Only 1 material with replicates for T other than room temperature, and only at one temperature

Park, Howe and Sholl, *Chemistry of Materials* 2017 (DOI: 10.1021/acs.chemmater.7b04287)
Recommendations

- Comprehensive meta-analysis can be a powerful tool
  
  *This requires compilation of comprehensive data sets*

- Meta-analysis doesn’t point out why some results are outliers or why observed range is wide/narrow

- Variability may arise from unreported (or unknown) details in materials synthesis and handling
  
  *Controlling these details is critical in real applications*

- Field-wide meta-analysis can create incentives to perform replicates
  
  *e.g., increase reproducibility of a materials by publishing targeted replicates*

- “Hidden” data may be a rich resource. Aligning incentives to reveal this data could greatly improve ability to assess reproducibility