

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

Chemical and Biomolecular Engineering **Georgia**Institute of **Tech**nology







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

Sholl and Lively, *Nature* 2016

Walton and Sholl, Joule 2017





Common characteristics in applied materials chemistry problems

- 1. Many (10³-10⁶) possible materials
- 2. Well defined metrics used to judge which material is "best"
- 3. Potential IP and commercial value create complications for data sharing

Georgialnstitute of Technology



Observation: In 'hot' topics, multiple groups study similar materials at similar times.



Image from history.com

Georgialnstitute of Technology

Implication: Replicate experiments may exist in literature even though replication was not the aim of the experiments.





Image from theskichannel.com

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

Georgialns

of **Tech**nology

• CO_2 adsorption in HKUST-1 at 298 \pm 5 K



18 replicates (27% classified as outliers)

Georgialnstitut of Technology





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







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

Georgialnsti

of **Tech**nology



- 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

Georgia Inst

of **Tech**nology