Poster

The Crystal Isometry Principle has made transparent the black boxes of structural databases

D. Widdowson[1], V. Kurlin[1]

[1]*Materials Innovation Factory and Computer Science Department, University of Liverpool, Liverpool, L69 3BX, UK*

vitaliy.kurlin@liverpool.ac.uk

After being developed over hundreds of years, in 2024 crystallography is witnessing a growing crisis of paper mills and artificial tools that generate hundreds of papers [1]. The rebuttals [2,3] of the two Nature papers [4,5] illustrate the following **common mistakes and problems in using black boxes**. The first paper [4] reported the GNoME database of 384+ thousand 'stable' predicted structures whose review [2] found '**scant evidence** for compounds that fulfill the trifecta of novelty, credibility, and utility'. The autonomous A- lab [5] claimed to have synthesized 43 new materials from the GNoME, whose review [3] concluded that 'none of the materials produced by A-lab were new: the large **majority were misclassified**, and a smaller number were correctly identified but **already known**'. The latest review [6] reported **thousands of (near-)duplicates** and even **1000+ pairs of identical CIFs** found as different entries in the GNoME.

The question "What material can be called new?" was recently resolved by generically invariants, which detected thousands of duplicates [7,8] in major experimental databases including the CSD, where HIFCAB and JEPLIA differ only by Cd swapped with Mn. The **Crystal Isometry Principle** says that all *real periodic crystals* live at *unique locations* in a *continuous space* whose simplest projection is below.

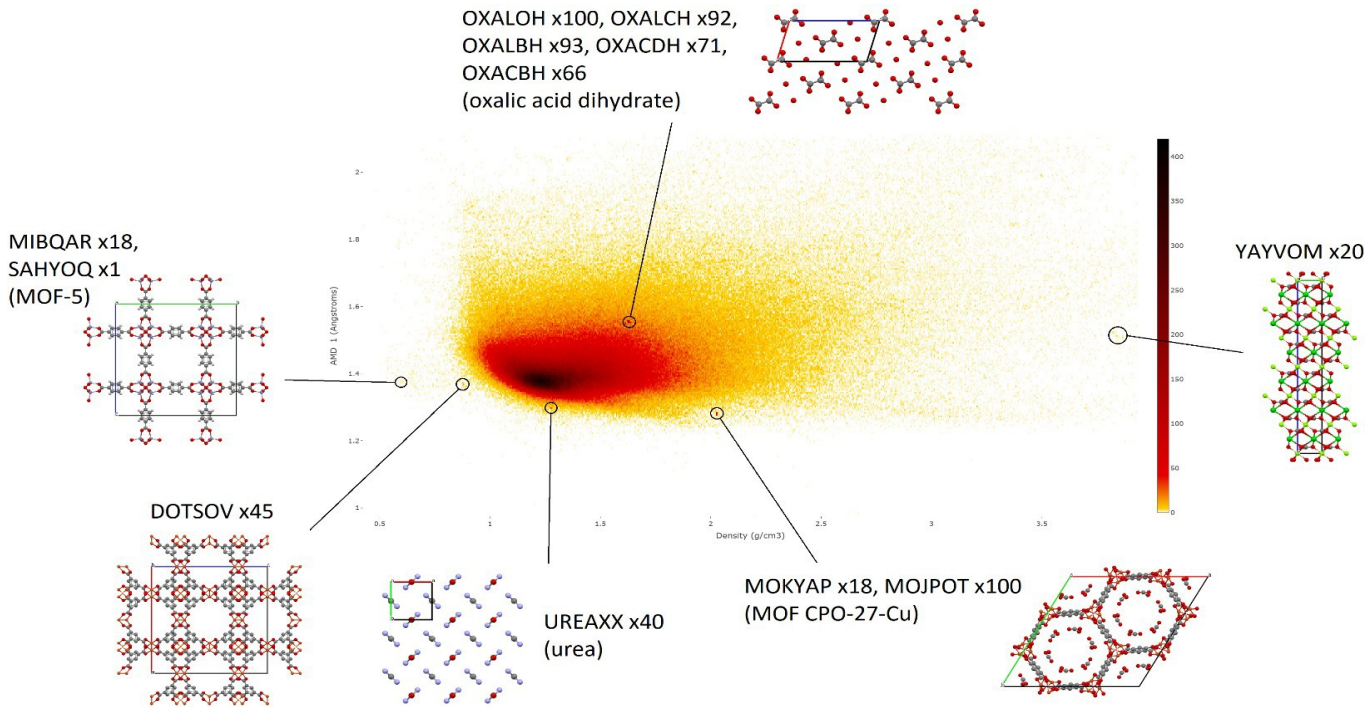

Figure 1. A continuous map of all 800K+ periodic crystals from the CSD: the density vs average distance to the first atomic neighbor.

[1] Chemical & Engineering News, https://cen.acs.org/research-integrity/Crystallography-databases-hunt-fraudulent-structures/102/i8, March 2024.
[2] Cheetham, A., Seshadri, R. Artificial Intelligence Driving Materials Discovery? Chemistry of Materials v. 36 (8), 3490–3495, April 2024.
[3] Leeman, J. et al. Challenges in High-Throughput Inorganic Materials Prediction and Autonomous Synthesis. PRX Energy 3, 011002. March 2024.
[4] Merchant, A. et al. Scaling deep learning for materials discovery. Nature v. 624 (7990), 80-85, November 2023.
[5] Szymanski, N. et al. An autonomous laboratory for the accelerated synthesis of novel materials. Nature v. 624 (7990), 86–91, November 2023.
[6] Anosova, O., Kurlin, V., Senechal, M. The importance of definitions in crystallography. IUCrJ, v.11(4), p.453-463, 2024.
[7] Widdowson, D., et al. Average Minimum Distances of periodic point sets. *MATCH Comm. Math. Comp. Chem*, v.87 (3), 529-559, 2022.
[8] Widdowson, D., Kurlin, V. Resolving the data ambiguity for periodic crystals. *Adv. Neural Inform. Proc. Systems*, v.35, 24625-24638, 2022.