From Molecular Mechanisms to Interpretable Machine Learning: Design Rules for Metal-Organic Frameworks in Atmospheric Water Harvesting

Nour Alkhatib[1,2], Saifeldeen Abed Alrhman[3], and Serdal Kirmizialtin[1,2,3]*

[1] Chemistry Program, Science Division, New York University Abu Dhabi, Abu Dhabi 129188, UAE
[2] Department of Chemistry, New York University, New York, NY 10003, USA
[3] Department of Chemical and Biomolecular Engineering, New York University Tandon School of Engineering, New York, USA
[4] Center for Smart Engineering Materials, New York University Abu Dhabi, Abu Dhabi 129188, UAE
email: na2416@nyu.edu

Atmospheric water harvesting with metal-organic frameworks (MOFs) could enable decentralized water production in arid regions, but selecting high-performing frameworks remains difficult because water uptake depends on the interplay among pore geometry, adsorption energetics, and local chemistry. Here, we combine molecular simulation with interpretable machine learning to identify the mechanisms and descriptors that control water adsorption from the level of individual pores to a chemically diverse materials database.

Atomistic simulations of benchmark aluminum MOFs show that adsorption begins when water anchors at polar framework sites and nucleates hydrogen-bonded clusters. The spatial distribution and chemical accessibility of these sites then control network percolation, the onset of uptake, and isotherm steepness. To test whether these molecular principles transfer across materials, we selected 2,600 structurally and chemically diverse frameworks from the ARC-MOF database and calculated water uptake at 30% and 100% relative humidity using high-throughput grand canonical Monte Carlo simulations. We trained and compared regression models using geometric, energetic, compositional, charge, and local chemical-environment descriptors.

Light Gradient Boosting Machine models achieved test R^2 values of 0.791 at 30% relative humidity and 0.848 at saturation. Removing energetic descriptors reduced the corresponding values to 0.679 and 0.720, demonstrating that pore geometry alone cannot explain water uptake. SHapley Additive exPlanations identified the Henry coefficient, localized oxygen and hydrogen charges, metal electronegativity, and framework density as the dominant predictors, while surface area and void fraction acted mainly as secondary modulators. A second-order polynomial surrogate built from the five highest-ranked descriptors provided an explicit relationship for rapid screening and hypothesis generation.

The work presented here establishes a multiscale framework that directly connects pore-level water-cluster formation with interpretable machine-learning predictions across 2,600 MOFs. Unlike conventional screening approaches that prioritize surface area and pore volume, our results quantitatively demonstrate that adsorption energetics and local electrostatics dominate water uptake, while geometry acts mainly as a secondary modulator. This connection yields a transferable design rule which is high-performing water-harvesting MOFs must combine accessible pore space with adsorption sites whose polarity and spatial arrangement initiate water clustering at low humidity without hindering regeneration. The workflow converts molecular mechanisms into chemically interpretable descriptors, providing a new route for the targeted, data-driven discovery of MOFs for atmospheric water harvesting.

Reference

[1] N. Alkhatib, S. A. Abed Alrhman, and S. Kirmizialtin, Mol. Syst. Des. Eng. 11, 50-61 (2026).