

Cite this: *J. Mater. Chem. A*, 2026, **14**, 10836

# Graph neural network-based multi-objective Bayesian optimization for enhanced screening of metal–organic frameworks with optimal separation performance

Lane E. Schultz,<sup>a</sup> Nickolas Gantzer,<sup>b</sup> N. Scott Bobbitt,<sup>b</sup> Dorina F. Sava Gallis<sup>c</sup> and Rémi Dingreville<sup>\*a</sup>

Metal–organic frameworks (MOFs) are porous crystalline materials with applications in gas capture, drug delivery, and molecular separations. While high-throughput computational screening has traditionally identified promising MOFs, recent advances increasingly harness machine learning to accelerate discovery and screening. Existing optimization methods such as Bayesian optimization (BO) and genetic algorithms often overlook the detailed structure–property relationships critical to MOF performance. Here, we present an optimization workflow that couples graph neural networks (GNNs) with multi-objective BO to enhance MOF discovery and screening. By representing MOFs as graphs embedding atomic- and structural-level features, GNNs capture intricate structure–property correlations, enabling more accurate property predictions than traditional methods relying solely on macroscopic descriptors. Our integrated framework efficiently identifies Pareto-optimal MOF candidates tailored for improved separation of alkanes, alkenes, alcohols, and aromatics, demonstrating the significant advantage of graph-based models in materials optimization workflows.

Received 10th November 2025  
Accepted 18th January 2026

DOI: 10.1039/d5ta09133k

rsc.li/materials-a

## 1 Introduction

Metal–organic frameworks (MOFs) are hybrid materials composed of two fundamental building blocks: (i) single metal ions or polynuclear metal clusters (secondary building units) and (ii) organic ligands that connect these units to form three-dimensional porous structures.<sup>1–3</sup> MOFs demonstrate exceptional utility across diverse applications,<sup>2</sup> including gas storage,<sup>4</sup> separations,<sup>5,6</sup> sensing,<sup>7,8</sup> luminescence,<sup>9,10</sup> environmental remediation,<sup>11</sup> catalysis,<sup>12,13</sup> and biomedical applications,<sup>14–16</sup> to name a few.<sup>8,17</sup> In sensing applications, the ability to selectively distinguish one analyte from others is crucial, and MOFs have demonstrated significant potential for volatile organic compound (VOC) detection through selective adsorption of specific target compounds over competing species.<sup>8</sup>

Due to their versatility, MOFs have garnered significant attention, with researchers actively seeking MOFs with enhanced properties for targeted applications. Given their modular construction and the vast chemical space of different

nodes, linkers, and topologies, there are a large number of possible MOFs, meaning they can be chosen or carefully designed for specific applications. Specialized databases containing thousands of MOFs have emerged, including the hMOF database,<sup>18</sup> ToBaCCo,<sup>19</sup> *ab initio* REPEAT charge MOF (ARC-MOF),<sup>20,21</sup> computation-ready experimental MOF (CoRE MOF),<sup>22</sup> and MOSAEC MOF database (MOSAEC-DB)<sup>23</sup> to name a few. ARC-MOF<sup>20</sup> features MOFs with density functional theory (DFT) calculated charges and some experimentally synthesized samples. CoRE MOF<sup>22</sup> stores experimentally synthesized MOFs along with some material properties. MOSAEC-DB<sup>23</sup> includes curated MOFs that were verified by the physical plausibility of metal oxidation states. These extensive databases provide researchers with opportunities for MOF property optimization on more physically feasible structures. While databases like CoRE MOF and ARC-MOF have aggregated tens of thousands of structures, they represent a vast and noisy chemical space. Brute-force high-throughput computational screening of every candidate for every possible application is computationally intractable. The generation of new MOFs has evolved dramatically from basic combinatorial approaches where ligands and secondary building units are interchanged between known structures,<sup>18</sup> to sophisticated machine-learning (ML) techniques such as variational autoencoders that explore latent spaces to produce innovative structures.<sup>24</sup> The primary bottleneck in materials discovery is no longer generation but rather

<sup>a</sup>Center for Integrated Nanotechnologies (CINT), Sandia National Laboratories, Albuquerque, NM 87185, USA. E-mail: rdingre@sandia.gov<sup>b</sup>Material, Physical, and Chemical Sciences Center, Sandia National Laboratories, Albuquerque, NM 87185, USA<sup>c</sup>Nanoscale Sciences Department, Sandia National Laboratories, Albuquerque, NM 87185, USA

the efficient screening and rapid identification of the highest-performing, Pareto-optimal candidates from these massive existing databases. Therefore, we utilize the well-established CoRE MOF<sup>22</sup> 2019 database for much of our work, which provides experimentally validated structures as a reliable foundation. We also use the ARC-MOF database.<sup>20,21</sup>

Various computational approaches have been employed to search for MOFs with optimal properties,<sup>25</sup> including hierarchical screening with ML predictions<sup>26</sup> and genetic algorithms.<sup>27</sup> Bayesian optimization (BO) has emerged as a particularly effective method, utilizing surrogate models and acquisition functions that balance exploitation (maximizing target properties if greater is better) and exploration (investigating uncertain areas) to efficiently identify promising candidates. Researchers have demonstrated that BO can rapidly screen optimal materials from large search spaces, sometimes requiring only a few proposed materials to find optimal candidates.<sup>28</sup> Advanced implementations incorporate multi-fidelity approaches that balance computational cost with accuracy for a similar class of material called covalent organic frameworks,<sup>29</sup> and diversity promoting mechanisms like the Vendi score to ensure broad exploration of a material space.<sup>30</sup> Others optimize multiple properties simultaneously like in Comlek *et al.*,<sup>31</sup> where MOFs with favorable CO<sub>2</sub> working capacity and CO<sub>2</sub> over N<sub>2</sub> selectivity were selected. Many approaches to MOF discovery employ BO with MOFs represented in a tabular format (*i.e.*, representations where each row corresponds to a material and each column to a feature or property). Tabular formats emphasize macroscopic properties such as pore size and surface area.<sup>32–34</sup> While useful, these simplified descriptors overlook critical structure–property relationships at the atomic scale inherent in graph-based representations. They fail to capture the local chemical environment of the metal node, the precise geometry of the organic linker, or the material's topology. Even strategies utilizing Simplified Molecular Input Line Entry System (SMILES) strings encoded as vectors, such as those discussed in Liu *et al.*,<sup>30</sup> may not inherently capture atomic properties like local chemistry or geometry. In contrast, graph neural networks (GNNs) that integrate both structural and chemical information have demonstrated superior performance in property prediction compared to their tabular counterparts, as highlighted by Luo *et al.*<sup>35</sup> and Li *et al.*<sup>36</sup> Both sources indicate that graph-based models achieve better accuracy in predicting properties than models trained on tabular data. Luo *et al.*<sup>35</sup> focused on atomic charge prediction while Li *et al.*<sup>36</sup> focused on prediction of CH<sub>4</sub> over N<sub>2</sub> selectivity. The integration of property optimization *via* BO with a GNN serving as the surrogate model holds significant promise for MOFs, offering a way to exploit complex structural and chemical information.

MOFs have been widely studied for chemical separations.<sup>5,6</sup> One of the most difficult challenges in adsorptive separations is to distinguish between molecules with similar chemical or geometric characteristics. When a MOF exhibits strong affinity for a specific molecule, it often shows comparable affinity for structurally or chemically related species, making selective adsorption difficult. Consequently, a tradeoff in selectivity inevitably arises, and this tradeoff becomes increasingly

pronounced as molecular similarity increases. In this work, we integrate a GNN-based structural representation of MOFs with a BO workflow to identify and screen MOFs capable of separating chemically similar organic molecules (*e.g.*, alcohols, butenes, benzene/toluene). This integrated approach enables the discovery of Pareto-optimal MOFs across multiple challenging separation tasks, effectively prioritizing candidates that optimally balance selectivity between closely related molecules.

We utilized structures from the ARC-MOF<sup>20,21</sup> v6 and CoRE MOF<sup>22</sup> 2019 Version 1.1.3 databases to search for Pareto-optimal MOFs. Each MOF was uniquely characterized using Voronoi polyhedra (VP), which defines the connectivity employed in a graph representation of the framework. These graphs incorporated atomic-level chemical information, local atomic neighborhood environments, and global MOF properties. Using these graph representations, surrogate GNNs were trained on subsets of MOFs and subsequently applied within a BO framework to explore the remaining search space, assessing whether the GNNs effectively identified Pareto-front materials. We validated this approach by comparing multiple model types with our GNN BO workflow to identify MOFs with optimal CH<sub>4</sub>/N<sub>2</sub> selectivity using the data of Li *et al.*<sup>36</sup> Separations were quantified by maximizing the Henry constant of the target species while minimizing those of competing molecules. The Henry coefficient, which represents the proportionality between gas adsorption and pressure in the low-pressure limit, was used as an indicator of affinity toward specific adsorbates. Optimization tasks included one, two, three, and four simultaneous chemical objectives. The GNN-based candidate selection consistently outperformed random selection across most test cases. Feature ablation analyses provided interpretability, revealing the respective contributions of edge-, node-, and global-level features to model predictions. The present framework not only captures structure-to-property mapping but also enables automated discovery of materials with optimized performance. The current search space, however, remains limited by the availability of structures within existing databases.

## 2 Computational methods

This section details the integrated computational workflow for screening MOF databases, generating the associated targeted Henry coefficients, and optimizing multi-property adsorption performance. This workflow consists of several elements. After curating structural datasets, Henry coefficients were computed as central descriptors for affinity and selectivity. MOF structures were featurized as graphs with rich node, edge, and global attributes. GNN-based surrogates were constructed and optimized to predict material properties and uncertainties, supported by feature ablation interpretation and benchmarking against Gaussian process (GP) surrogates. These surrogate models were iteratively coupled with BO to efficiently identify MOFs with optimal adsorption properties, using acquisition functions and performance metrics for both single- and multi-objective searches. The sub-sections below describe each of these steps in more detail.



## 2.1 MOF database

We used a subset of 77 482 and 18 522 MOF structures from the ARC-MOF<sup>20,21</sup> v6 and CoRE MOF 2019 databases,<sup>22,37</sup> respectively. The ARC-MOF collection includes both experimentally synthesized and computationally generated frameworks, along with DFT-computed atomic charges and several macroscopic MOF properties. The CoRE MOF database comprises experimentally realized MOFs that have been curated for atomistic modeling by addressing issues such as partial occupancies, symmetry inconsistencies, and the presence of solvent molecules. The structural information for each MOF in both datasets is encoded in a crystallographic information file (CIF) containing the element type, site indices, and three-dimensional atomic spatial coordinates. From CoRE MOF, subsets were further filtered to only retain structures with a pore limiting diameter greater than the largest kinetic diameter of adsorbed VOCs. See the SI for tabulated kinetic diameters. We note that these datasets primarily comprise ‘computationally ready’ versions of known structures. While such readiness increases the likelihood that our ‘optimal’ materials can be synthesized, it also means that the model may have lower predictive confidence in ‘dark’ chemical spaces, such as those involving highly reactive open-metal sites or non-traditional coordination geometries that are underrepresented in the training data.

## 2.2 Henry constants and the setup for optimization

The adsorption affinity of a gas in the dilute limit can be described by the Henry coefficient which quantifies the proportionality between the amount adsorbed and its partial pressure. This metric offers a convenient and computationally efficient means to evaluate host–guest interactions under infinite dilution, where adsorbate–adsorbate interactions are negligible. Because Henry coefficients capture intrinsic framework–adsorbate affinities, they serve as an effective descriptor for screening materials in early-stage separation studies. Mathematically, Henry’s law defines the coefficient as,

$$H = \frac{q}{P}, \quad (1)$$

where  $q$  is the amount adsorbed per unit mass of adsorbent and  $P$  is the partial pressure. Larger  $H$  values indicate stronger adsorption affinity. Selectivity for binary separations can be calculated as the ratio of Henry coefficients in the dilute limit:<sup>38</sup>

$$S_{1,2} = \frac{H_1}{H_2}, \quad (2)$$

where  $S_{1,2}$  represents the preferential selectivity of component 1 over component 2. For multi-component systems where one component is preferentially adsorbed while others are excluded, selectivity is maximized by increasing the Henry coefficient of the desired species while minimizing those of competing ones. For example, given three adsorbates with Henry coefficients  $H_1$ ,  $H_2$ , and  $H_3$ , maximizing  $H_1$  while minimizing  $H_2$  and  $H_3$  identifies materials most favorable for adsorbing component 1. This optimization logic extends to selecting any subset of materials over competing species by appropriately maximizing desired Henry coefficients while

minimizing those of excluded components. We note that there are cases where a MOF has different adsorption sites which show preference for different adsorbates, but that is rare and not covered in this work.

Finding a single material that optimizes multiple selectivities is often infeasible due to inherent trade-offs in material properties. Instead, a set of non-dominated solutions, known as the Pareto front, can be identified for multi-objective optimization. The Pareto front represents solutions that cannot be improved in one objective without degrading another.<sup>39</sup> Mathematically, a solution  $y_i$  belongs to the Pareto front if no other solution  $y_j$  dominates it:

$$\forall y_i \in P : \nexists y_j \in S : y_j \succ y_i, \quad (3)$$

where  $P$  is the Pareto front,  $S$  is the solution space, and  $y_j \succ y_i$  indicates that  $y_j$  dominates  $y_i$  (*i.e.*,  $y_j$  is superior in at least one objective and at least equivalent if not better in all others). For our selectivity problem, this corresponds to materials that best balance competing selectivities such as  $S_{1,2} = H_1/H_2$  and  $S_{1,3} = H_1/H_3$ . Because chemically similar adsorbates often produce strongly correlated Henry coefficients, trade-offs in selectivity are expected and must be captured explicitly through Pareto analysis. The dataset, number of structures, target property or objective, functional group family, and optimization direction (*i.e.*, maximization or minimization) explored in this study are summarized in Table 1.

## 2.3 Data generation for volatile organic compounds Henry coefficients

Henry coefficient data were generated using Widom insertions simulations implemented in RASPA<sup>40</sup> v2, closely following the methods described in Bobbitt *et al.*<sup>41</sup> RASPA is an open-source molecular simulation software designed for adsorption and diffusion studies in molecules such as MOFs. Each simulation was performed for 20 000 insertion cycles per adsorbate. MOF structures were held rigid at their crystallographic minimum positions. Non-bonded interactions were treated using 12–6 Lennard-Jones potentials with parameters taken from the universal force field (UFF)<sup>42</sup> for MOFs and TraPPE<sup>43–46</sup> for the adsorbates. For most adsorbates, the united atom model was applied, while for benzene and toluene we employed rigid 12-site models with explicit charges. The toluene model was adapted from benzene by replacing a hydrogen atom with an  $sp^3$  methyl group. Interaction cross terms were determined using Lorentz–Berthelot mixing rules. Partial charges on the MOFs were computed using the EQeq<sup>47</sup> method as implemented in RASPA and treated using Ewald summations. All simulations used a 12.0 Å cutoff radius and 300 K temperature, with simulation boxes duplicated as needed to ensure dimensions at least twice this value in all directions. Ideal gas Rosebluth weights were pre-computed at 300 K. Given the wide range of values in Henry coefficients, we took their natural logarithms for subsequent analysis. All Lennard-Jones parameters and molecule structures are provided as RASPA definition files in the repository mentioned in the Data availability section at the end of this article.



**Table 1** Datasets for multi-objective Bayesian optimization (MOBO) with their corresponding total number of studied structures, objectives, kind of functional group family, and whether they are maximized (Max) or minimized (Min) in optimization studies

Dataset	Structures	Objective	Family	Min/Max
{1-propanol}	3634	1-Propanol	Alcohol	Min
{Butane}	3653	Butane	Alkane	Max
{Benzene, toluene}	2125	Benzene	Aromatic	Max
		Toluene	Aromatic	Min
{1-propanol, butane}	3634	1-propanol	Alcohol	Min
		Butane	Alkane	Max
{1-propanol, 2-propanol}	3634	1-propanol	Alcohol	Max
		2-propanol	Alcohol	Min
{1-propanol, propane}	3634	1-propanol	Alcohol	Min
		Propane	Alkane	Max
{2-cis-butene, 2-trans-butene}	3260	2-cis-butene	Alkene	Max
		2-trans-butene	Alkene	Min
		Butane	Alkane	Max
{Butane, propane}	3653	Butane	Alkane	Max
		Propane	Alkane	Min
{1-propanol, 2-propanol, butane}	3634	1-propanol	Alcohol	Min
		2-propanol	Alcohol	Min
{1-propanol, 2-propanol, propane}	3634	Butane	Alkane	Max
		1-propanol	Alcohol	Min
		2-propanol	Alcohol	Min
{1-propanol, 2-propanol, butane, propane}	3634	Propane	Alkane	Max
		1-propanol	Alcohol	Min
		2-propanol	Alcohol	Min
		Butane	Alkane	Max
		Propane	Alkane	Max

Regarding the representativeness of Henry coefficients as indicators of separation performance, we note that Henry's law strictly applies only at low adsorbate concentrations. Nevertheless, Henry coefficients remain informative for low-concentration applications such as VOC capture or chemical sensing.<sup>48,49</sup> They are also valuable descriptors for dilute gas separations, consistent with prior studies.<sup>50</sup> Moreover, Henry coefficients computed *via* Widom insertions capture a comprehensive average of adsorbate–framework interactions by sampling the entire pore environment. Accounting for higher-concentration behavior would require explicitly modeling adsorbate–adsorbate interactions, which become non-negligible for species such as alcohols. Incorporating these effects would entail substantial methodological extensions that we identify as an important avenue for future work.

## 2.4 MOF featurization

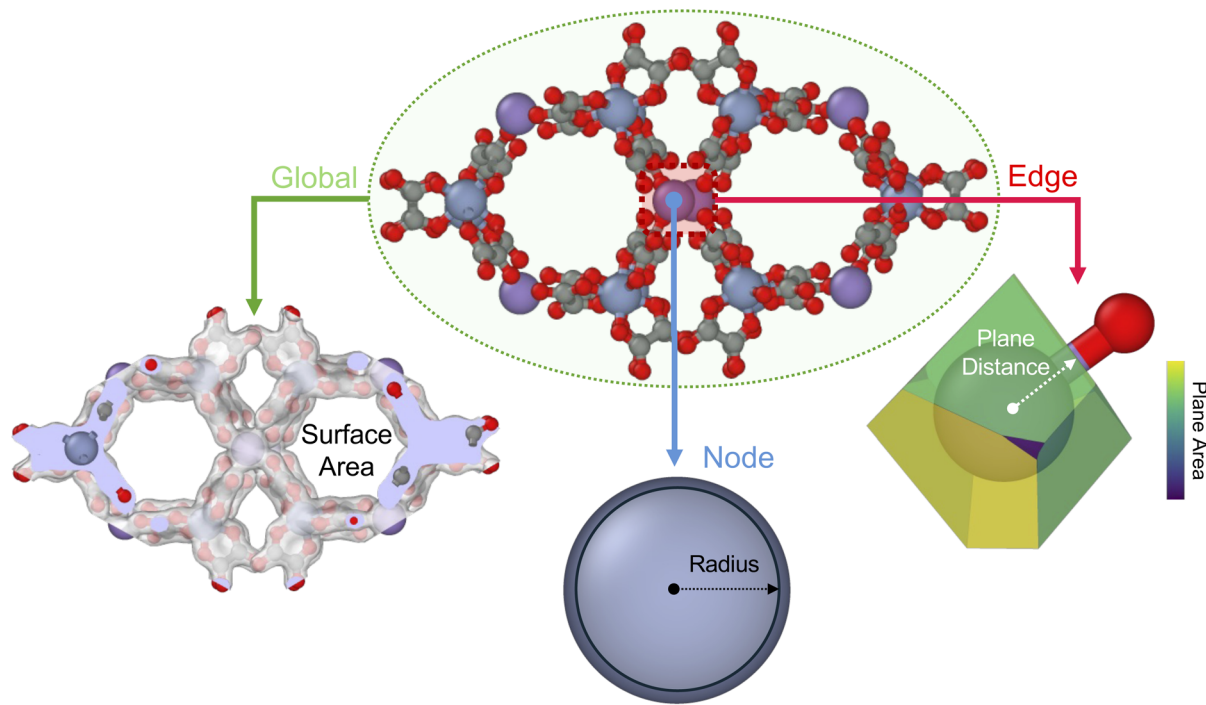
**2.4.1 Graph representation.** Before screening for MOFs with optimized properties, each structure must first be represented as a graph so that a GNN can uniquely map MOF structural information to property predictions. A graph consists of sets of nodes and edges. Each edge denotes a connection between pairs of nodes. Features can be encoded at multiple levels: within individual nodes, along edges, and across the graph as a whole like shown in Fig. 1(a). In this representation, atoms are treated as nodes, and their connections to nearest neighbors serve as edges (see panels (b) and (c) in Fig. 1). This graph-based approach provides a systematic and compact way to encode the complex three-dimensional topology of MOFs. Importantly, these representations are invariant to permutation (*i.e.*, shuffling site order), translation, and rotation, ensuring

that a given MOF maintains a unique representation regardless of spatial orientation. Throughout this work, the term node refers exclusively to its graph-theoretic definition rather than a structural unit of the MOF.

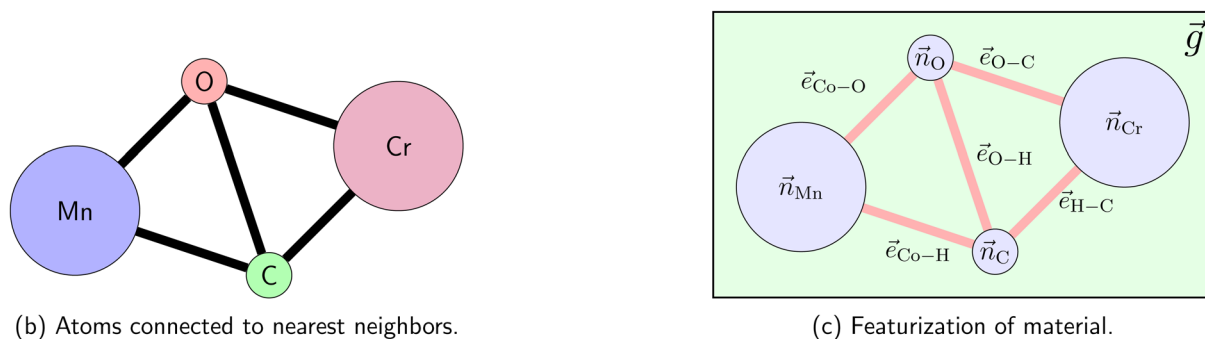
The connections between nodes forming the graph structure were defined using VP, a geometric construction that partitions space around each atom by defining bisecting planes between the central atom and its neighboring atoms. Each MOF structure provided as a CIF was loaded into pymatgen<sup>51</sup> as a structural object. The nearest-neighbor network was constructed using pymatgen's *IsayevNN* class, applying a bond distance tolerance of 0.5 Å without imposing restrictions on the number of Voronoi neighbors.<sup>51,52</sup> This strategy, inspired by Luo *et al.*,<sup>35</sup> defines an edge between two atoms if (i) their distance is less than or equal to the sum of their Cordero covalent radii plus a distance tolerance and (ii) they share a common Voronoi facet. The 0.5 Å tolerance was specifically chosen to prevent isolated nodes, as noted in Luo *et al.*<sup>35</sup> Thus each MOF was represented as a graph in which atomic connectivity is encoded through VP-based spatial relationships.

Other featurizations exist. We acknowledge that the smooth overlap of atomic positions (SOAP)<sup>53</sup> approach is a powerful method for encoding local atomic information in structures. SOAP representations are differentiable with respect to atomic positions (crucial for molecular dynamics simulations) while maintaining the invariance properties characteristic of graph representations. However, the original SOAP formulation incorporates only positional and element type information and does not incorporate additional chemical descriptors such as electronegativity. In contrast, graph representations offer





(a) Example of MOF features.



(b) Atoms connected to nearest neighbors.

(c) Featurization of material.

$$X = \left\{ \begin{array}{l} \left[ \begin{array}{c} g_1 \\ g_2 \\ \vdots \\ g_x \end{array} \right]^T, \left[ \begin{array}{cccc} n_{\text{Mn},1} & n_{\text{Mn},2} & \cdots & n_{\text{Mn},y} \\ n_{\text{O},1} & n_{\text{O},2} & \cdots & n_{\text{O},y} \\ n_{\text{C},1} & n_{\text{C},2} & \cdots & n_{\text{C},y} \\ n_{\text{Cr},1} & n_{\text{Cr},2} & \cdots & n_{\text{Cr},y} \end{array} \right], \left[ \begin{array}{cccc} e_{\text{Mn-O},1} & e_{\text{Mn-O},2} & \cdots & e_{\text{Mn-O},z} \\ e_{\text{Mn-C},1} & e_{\text{Mn-C},2} & \cdots & e_{\text{Mn-C},z} \\ e_{\text{O-C},1} & e_{\text{O-C},2} & \cdots & e_{\text{O-C},z} \\ e_{\text{O-Cr},1} & e_{\text{O-Cr},2} & \cdots & e_{\text{O-Cr},z} \\ e_{\text{C-Cr},1} & e_{\text{C-Cr},2} & \cdots & e_{\text{C-Cr},z} \end{array} \right] \end{array} \right\}$$

(d) Set of graph features.

Fig. 1 Conceptual framework for graph-based representation of metal-organic frameworks (MOFs). A MOF from CoRE MOF<sup>22,37</sup> is shown in panel (a) along with a subset of calculated properties. Panel (b) presents an abstraction of a material structure. Panel (c) illustrates the corresponding graph representation, where atoms are represented as nodes and connections to neighbors as edges. Panel (d) depicts the global features that characterize the entire graph, the node features that encode atomic properties, and the edge features that capture connectivity.

greater flexibility, allowing the seamless inclusion of diverse chemical and structural features within a unified framework.

**2.4.2 Edge, node, and global features.** Edge features were derived from the VP analysis to capture local coordination and geometric information critical for differentiating atomic environments within the MOF structure. For each pair of connected atoms, four quantities were encoded: the distance from the central atom to the bisecting plane, the plane area, the number

of vertices defining the plane, and the fractional volume contribution of the facet to the total atomic enclosure.

Chemical and structural properties of each atomic site were also incorporated as node features. These properties were obtained from two sources: descriptors extracted through pymatgen<sup>51</sup> and features adopted from Luo *et al.*<sup>35</sup> From pymatgen, each node was encoded with elemental attributes including Mendeleev number, atomic mass, number of electrons, electron



affinity, and the number of electrons in each orbital of the atom up to the 7p orbital (with unoccupied orbitals assigned a value of zero). The VP geometry of each atom was represented using Schläfli notation, which records the specific number of enclosing faces and the number of vertices per face. For example, an atom with nine enclosing faces, where three faces have four vertices and the remaining six have five vertices, is represented as (0, 3, 6), where the index indicates vertex count and is the number of faces. Faces with up to 10 vertices were considered, following the hard-coded constraint in pymatgen's VoronoiAnalyzer. In total, 31 node features were obtained from the properties described above, while additional features were included using modified code and parameterizations provided by Luo *et al.*<sup>35</sup>

As in Luo *et al.*,<sup>35</sup> each node was further enriched with atomic properties from Mendeleev,<sup>54</sup> including atomic number, atomic weight, van der Waals radius, first ionization energy, Cordero covalent radius, Rahm atomic radius, static dipole polarizability, and Ghosh electronegativity (8 features). Beyond these direct attributes, a set of local environment descriptors were computed to capture geometric and chemical interactions between neighboring atoms.

The property weighted radial distribution functions (pwRDFs)<sup>55</sup> quantify how a given atomic property  $P$  varies radially around each atom  $i$ :

$$\text{pwRDF}_i(R) = \sum_{i \neq j}^N P_j \cdot e^{-\alpha(r_{ij}-R)^2}, \quad (4)$$

where  $r_{ij}$  is the distance between atom  $j$  and atom  $i$ ,  $N$  is the total number of atoms surrounding  $i$ ,  $R$  is the distance of interest from atom  $i$ , and  $\alpha$  is a Gaussian smoothing parameter. Only atoms within a 8 Å cutoff distance of atom  $i$  were considered, yielding a total of 28 radial features.

The property weighted angular distribution functions (pwADFs) were computed analogously to describe three-body correlations between atom  $i$  connected to atoms  $j$  and  $k$ :

$$\text{pwADF}_i(\Theta) = \sum_{i \neq j}^N \sum_{k \neq i, j}^N P_j \cdot P_k \cdot e^{-\beta(\theta_{ijk}-\Theta)^2}, \quad (5)$$

where  $\theta_{ijk}$  is the bond angle centered on atom  $i$ ,  $\beta$  is a Gaussian smoothing parameter, and  $\Theta$  is a given angle of interest. For this descriptor, we used a 6 Å cutoff distance, yielding 18 features. Normalized forms of these two descriptors, the radial weighted average atomic properties (RWAAPs) (eqn (6)) and angular weighted average atomic properties (AWAAPs) (eqn (7)), follow the same notation and yielded 14 and 16 additional features, respectively.

$$\text{RWAAP}_i(R) = \frac{\text{pwRDF}_i(R)}{\sum_{i \neq j}^N e^{-\alpha(r_{ij}-R)^2}}, \quad (6)$$

$$\text{AWAAP}_i(\Theta) = \frac{\text{pwADF}_i(\Theta)}{\sum_{i \neq j}^N \sum_{k \neq i, j}^N e^{-\beta(\theta_{ijk}-\Theta)^2}}. \quad (7)$$

The weighed radial atom centered symmetry function (wRACSF)<sup>35,56,57</sup> was also included,

$$\text{wRACSF}_i(R) = \sum_{i \neq j}^N P_j \cdot e^{-\gamma(r_{ij}-R)^2} \cdot f_{\text{cut}}(r_{ij}), \quad (8)$$

with the cutoff function

$$f_{\text{cut}}(r_{ij}) = \frac{1}{2} \left[ \cos\left(\frac{\pi r_{ij}}{R_{\text{cut}}}\right) + 1 \right], \quad (9)$$

producing 28 features. The wRACSF shares conceptual similarities with pWRDF, but introduces two key distinctions: the implementation of a cutoff function (eqn (9)) and the utilization of a different Gaussian smoothing parameter,  $\gamma$ , which serves to differentiate these two descriptor functions. The cutoff function emphasizes short range features within some distance  $R_{\text{cut}}$ . Similarly, weighed angular atom centered symmetry function (wAACSF)<sup>35,57</sup> added angle-dependent terms and phase modulation parameter  $\lambda = \pm 1$ , and another Gaussian smoothing parameter noted as  $\eta$ , generating 18 features.

$$\text{wAACSF}_i(\Theta) = \sum_{i \neq j}^N \sum_{k \neq i, j}^N P_j \cdot P_k \cdot (1 + \lambda \cdot \cos(\theta_{ijk})) \cdot e^{-\eta(r_{ij}+r_{ik}+r_{jk})^2} \cdot f_{ij} \cdot f_{ik} \cdot f_{jk}. \quad (10)$$

Finally, shell features<sup>35,58</sup> characterized property distributions over coordination shells up to the fourth order. We first define what a coordination shell for an atom  $i$  represents. The first coordination shell consists of the immediate neighbors surrounding an atom. The second coordination shell encompasses all atoms connected to the first neighbors of atom  $i$ . The third coordination shell includes all atoms connected to the second coordination shell atoms, and this pattern continues for higher-order shells. For each atom in a MOF and for each property  $P$  and shell  $n$ , we computed the average, absolute difference, and standard deviation:

$$\text{Shell}_p^{\text{avg}}(i) = \sum_i^N \frac{P_j}{N}, \quad (11)$$

$$\text{Shell}_p^{\text{dif}}(i) = \sum_i^N \frac{|P_i - P_j|}{N}, \quad (12)$$

$$\text{Shell}_p^{\text{std}}(i) = \sqrt{\sum_i^N \frac{(P_i - P_j)^2}{N}}. \quad (13)$$

producing 96 additional features (corresponding to 8 atomic properties, 4 shells, and 3 statistical operators).

Global features were classified into two categories: material-level and graph-based descriptors. Material-level features were computed from structural and pore geometry data, while graph-based features were derived exclusively from the topological representation of each MOF. Material-level properties, calculated using pymatgen,<sup>51</sup> include volume, density, the number of atoms, the number of each element type in a structure normalized by the total number of atoms, the average VP



volume for all atoms, and the average number of atom neighbors. Additional geometrical descriptors evaluated with Zeo++,<sup>59</sup> include the diameter of the largest included sphere, diameter of largest free sphere, diameter of largest included sphere along the free sphere path, both available and unavailable surface area, and both the available and unavailable void fraction. A spherical probe radius of N<sub>2</sub> was used. Graph-based descriptors, computed using NetworkX,<sup>60</sup> describe structural connectivity and topology. These include undirected graph density, diameter, degree assortativity coefficient, and average shortest path length. The graph density is defined as  $\frac{2e}{n(n-1)}$  where  $e$  and  $n$  are the number of edges and nodes respectively and describes the ratio of actual connections to possible connections in the graph. The graph diameter and radius correspond to the maximum and minimum eccentricities across all nodes, representing the farthest and nearest network spans, respectively. Unconnected nodes were excluded from distance calculations. The degree assortativity coefficient is the Pearson correlation coefficient of the degree between pairs of linked nodes, while the average shortest path length is given by

$$l = \sum_{\substack{i,j \in N \\ i \neq j}} d(i,j) = \frac{d(i,j)}{n(n-1)}, \quad (14)$$

where  $d(i,j)$  is the path distance between nodes  $i$  and  $j$ ,  $n$  is the number of nodes, and  $N$  is the set of nodes in a graph. In total, 135 global features were obtained across all materials and graph categories.

## 2.5 Construction, optimization, and feature interpretation of graph neural networks for multi-property regression

Efficient screening of a large set of MOFs with optimized performance involves evaluating thousands of candidates across multiple objectives, which is computationally demanding. BO provides a data-efficient framework for this purpose by iteratively selecting new candidates based on model predictions and associated uncertainties. To perform BO, a surrogate model capable of predicting target properties and quantifying its own uncertainty is essential. While GPs naturally satisfy these requirements, they are computationally expensive and scale poorly with data size. To overcome these limitations, we developed a neural surrogate model based on leveraging the flexible nature of neural networks (NNs), which can learn non-linear structure–property relationships while estimating prediction uncertainties. This framework combines the expressive power of NNs with the uncertainty awareness critical for BO-driven materials screening. Our implementation of GNNs utilizes the PyTorch Geometric (PyG)<sup>61</sup> library.

The GNN adopted here follows the formulation introduced by Battaglia *et al.*<sup>62</sup> and implemented through the MetaLayer class in PyG. In this architecture, edge, node, and global sub-models learn (from features scaled with scikit-learn's RobustScaler<sup>63</sup>) complementary hierarchical representations that capture multi-scale structural information within graphs. The latent features extracted from these components are subsequently processed through additional neural layers to infer material properties and

their uncertainties. Fig. 2 provides an abstraction of our GNN, showing how multi-level graph features (global, node, and edge) are encoded into latent representations for predicting target properties with uncertainties.

**2.5.1 Latent structure representations and incorporation into a graph neural network.** To accurately capture the hierarchical nature of structural information in MOFs, the GNN was designed with three interdependent sub-models: edge, node, and global. Each sub-model is responsible for learning representations at a different structural level of the graph. The edge sub-models capture local pairwise interactions between connected atoms, the node sub-models incorporate neighborhood information aggregated from these local interactions, and the global sub-model integrates system-wide features reflecting overall material characteristics. This hierarchical decomposition allows the network to combine information across scales, from atomic coordination environments to global framework topology, producing a comprehensive latent representation suitable for multi-property prediction. The edge sub-model constructs latent edge features by combining information from both the source and destination nodes, along with associated edge and global features. These combined features are processed through a NN, producing a latent edge representation that captures pairwise geometric and chemical interactions. The update function  $\phi_e$  for an edge,  $\vec{e}$ , connecting nodes  $\vec{n}_i$  and  $\vec{n}_j$ , given a global feature vector,  $\vec{g}$ , is defined as,

$$\vec{c}_{i,j} = \phi_e(\vec{n}_i, \vec{n}_j, \vec{e}_{i,j}, \vec{g}), \quad (15)$$

where  $\vec{c}$  denotes the new representation of edges. The input layer dimensionality of  $\phi_e$  is equal to twice the number of node features (comprising features from source and destination nodes), plus the number of edge features and global features. All hidden layers and the output layer maintain consistent width, with rectified linear unit (ReLU) activations and batch normalization between layers. The number of layers and neurons were later optimized, as described in Section 2.5.3.

The node sub-model builds upon latent edge representations to encode neighborhood effects. For each node, latent edge vectors directed toward it are aggregated using the PyG scatter function, combined with its intrinsic node and global features, and passed through a neural network  $\phi_n$ :

$$\vec{m}_i = \phi_n(\vec{n}_i, \rho^{e \rightarrow n}(\vec{c}_{j,i}), \vec{g}) \quad (16)$$

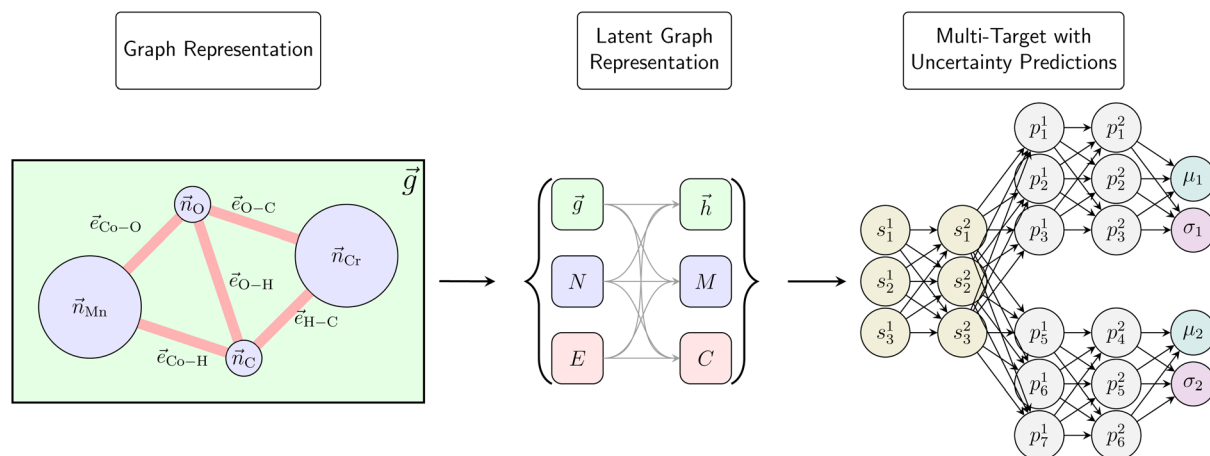
where  $\rho^{e \rightarrow n}$  denotes the aggregation operator and  $\vec{m}$  is the latent representation of the nodes. This produces latent node features that reflect both local geometry and broader contextual information from neighboring atoms. The architecture of  $\phi_n$  follows the same design as  $\phi_e$ .

The global sub-model integrates information across the entire structure, combining aggregated node and edge embeddings with global descriptors  $\vec{h}$ . Its update function is given by

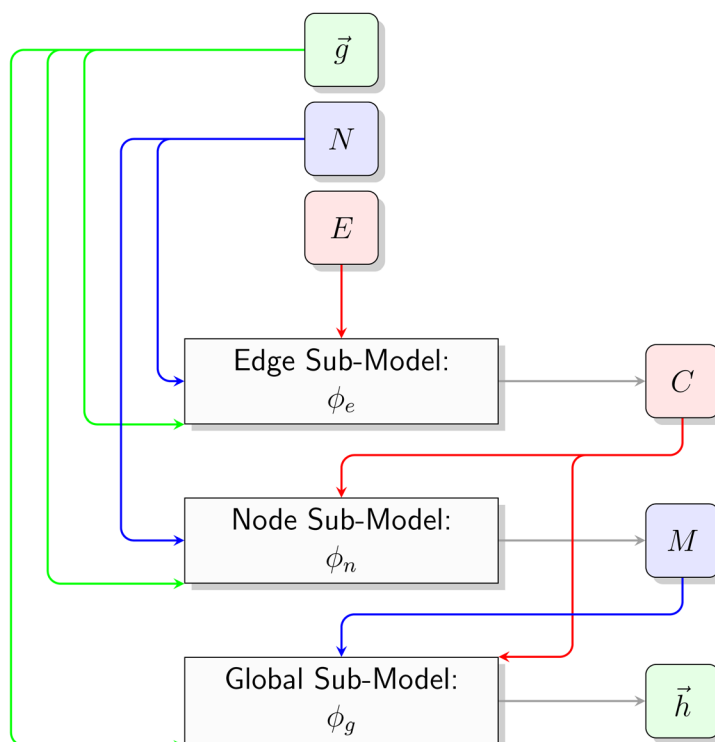
$$\vec{h} = \phi_g(\vec{g}, \rho^{n \rightarrow g}(\vec{m}_i), \rho^{e \rightarrow g}(\vec{c}_{j,i})) \quad (17)$$

where  $\rho^{n \rightarrow g}(\vec{m}_i)$  and  $\rho^{e \rightarrow g}$  perform mean aggregations across all nodes and edges, respectively. The network  $\phi_g$  adheres to the same architectural conventions as  $\phi_e$  and  $\phi_n$  sub-models.





(a) Inference from graph by surrogate model.



(b) Latent graph representation.

**Fig. 2** Graph-neural network (GNN) architecture for the surrogate. Panel (a) illustrates the computational workflow of a GNN surrogate model, which accepts a molecular graph as input and propagates structural features through a latent embedding space. The architecture consists of shared layers that process the latent representation, followed by task-specific separated layers that generate both predicted target quantities and associated uncertainty estimates. Panel (b) demonstrates the feature aggregation mechanism, showing how edge, node, and global features are combined to construct the latent representation.

After the latent edge, node, and global representations were established, their outputs were concatenated and passed through a fully connected NN that produces final material-level embeddings (the  $s$  neurons in Fig. 2(a)). These embeddings were then connected to separate layers corresponding to each target property (the  $p$  neurons in Fig. 2(a)). Each output head contains twice the number of neurons as target properties—one

neuron for the predicted value and another for the associated uncertainty—allowing the model to learn both quantities simultaneously.

**2.5.2 Learning target properties and uncertainties.** Pre-processing of input features and target variables was essential for stable model training. Edge-, node-, and global-level features, together with target values, were concatenated across



a training set, and reduced such that their variance was above zero (*i.e.*, no constant values). All remaining variables were scaled to have zero median and unit interquartile range (*i.e.*, the range between the first and third quartiles) to reduce the influence of outliers. During inference, the same scaling parameters were applied to inputs, while predicted targets were inversely transformed to recover their original physical scale.

Model training employed the negative log-likelihood (NLLH) loss function. We assumed Gaussian-distributed prediction errors. For each sample  $i$  and target property  $y$ , the model generated a mean prediction  $\mu_i$  and an associated standard deviation  $\sigma_i$ . The loss, later averaged over all training instances and target properties, is expressed as

$$L(y_i, \mu_i, \sigma_i) = \frac{1}{2} \log(2\pi) + \log(\sigma_i) + \frac{(y_i - \mu_i)^2}{2\sigma_i^2}. \quad (18)$$

While all Henry coefficient targets were equally weighted here, the framework allows the use of property-specific weights to emphasize selected objectives during training.

**2.5.3 Architecture optimization and extended training.** We performed hyperparameter optimization of the GNN architecture using Optuna,<sup>64</sup> adjusting the number of hidden layers and neurons per layer. The hyperparameter search space allowed between one and five hidden layers, with each layer containing between one and 1000 nodes. Graphs were divided into training, validation, and test sets with a 80 : 20 : 0 ratio for BO and a 80 : 10 : 10 ratio for parity plots, respectively. Fixed training parameters included a learning rate of 0.001, a batch size of 32, an early stopping criterion of 10 epochs (*i.e.*, the number of epochs with no improvement to terminate learning), and an exponential learning rate scheduler with  $\gamma$  of 0.99. Each of the 100 Optuna trials involved a short 10-epoch training run. Model performance was assessed using the area under the average NLLH loss function *versus* epoch curve, normalized by the number of epochs to mitigate the effects of early stopping. This approach provided a smoother and more reliable optimization metric than using the minimum or final epoch loss. We selected the optimal architectural parameters from these trials. The final model parameters were taken from the epoch with the lowest observed loss. For BO, this full optimization and training workflow was repeated at each iteration to refine the surrogate model continuously.

**2.5.4 Feature importance of optimal graphs.** Feature importance at the node, edge, and global levels was interpreted for an individual MOF using feature ablation, where predictions for each target property were compared against those from ablated graphs with features set to zero (the median from scaling). Correlated effects occur when feature ablation decreases (increases) the average feature value while decreasing (increasing) the GNN output with respect to the original predictions. The inverse correlation occurs when one increases as the other decreases. The magnitude of change in predicted properties quantified the influence of that feature on the GNN predictions. Feature importances were normalized to sum to one, and the ten most influential features were reported for comparison. Although variance-based methods such as Sobol sensitivity analysis can measure coupled feature effects, their computational cost scales poorly with the large number of input variables

(135 global, 96 node, and 4 edge features). Feature ablation offers a more efficient and interpretable alternative that highlights the substantial contribution of microscopic (node- and edge-level) features alongside macroscopic (global-level) descriptors.

## 2.6 Gaussian process as a surrogate model

We compared our GNN-based surrogate with a commonly used surrogate model: a GP. GP is a non-parametric model that treats a set of points as jointly following a multivariate Gaussian distribution. By conditioning this distribution on observed data, the GP yields predictive means and standard deviations for new inputs. We implemented the SingleTaskGP from BoTorch<sup>65</sup> with ExactMarginalLogLikelihood to train surrogate GP models, following the approach used in Deshwal *et al.*<sup>28</sup> Our GP-based BO workflow was validated against their published results, with corresponding verification details provided in the SI. For comparisons with more than one objective, the KroneckerMultiTaskGP model type was used. Because GP models operate on fixed-length tabular inputs, they were trained solely on global-level features extracted from the graph representations of MOFs.

## 2.7 Bayesian optimization

Given that both the GNN- and GP-based surrogates provide mean predictions and associated uncertainties, they can be directly applied within BO. Algorithm 1 outlines the general procedure. The process begins by fitting a surrogate model capable of estimating property values and uncertainties. Candidate materials are evaluated using an acquisition function that ranks them according to their expected improvement or contribution to the Pareto front. The top-ranked candidate is then selected, evaluated for its true properties, and appended to the training dataset. The surrogate is retrained on the expanded dataset, and the cycle repeats until a predetermined budget is reached. The budget may correspond to the number of BO iterations (as used in this study), total computation time, or other practical constraints.

---

### Algorithm 1 Bayesian optimization

---

**Input:** Initial dataset  $\mathcal{D}_0 = \{(x_i, y_i)\}_{i=1}^{n_0}$ , surrogate model  $\mathcal{M} : \mathbf{x} \mapsto N(\hat{y}(\mathbf{x}), \sigma^2(\mathbf{x}))$ , acquisition function  $\alpha$ , maximum number of iterations  $T$

Fit the surrogate model  $\mathcal{M}$  to the initial dataset  $\mathcal{D}_0$

**for**  $t = 1$  to  $T$  **do**

Optimize the acquisition function  $\alpha$  to select the next point to sample:

$$x_{t+1} = \arg \max_{x \in X} \alpha(x; \mathcal{M})$$

Evaluate the objective function at  $x_{t+1}$  to obtain  $y_{t+1}$

Augment the dataset with the new observation:

$$\mathcal{D}_{t+1} = \mathcal{D}_t \cup \{(x_{t+1}, y_{t+1})\}$$

Update the surrogate model  $\mathcal{M}$  with the new dataset  $\mathcal{D}_{t+1}$

**end for**

---

We applied BO to explore the space of MOFs with optimal Henry coefficients. Each surrogate was initially trained with 25 data points. For single-objective Bayesian optimization (SOBO), we used the expected improvement (EI) acquisition function.



For multi-objective Bayesian optimization (MOBO), we used the expected hypervolume improvement (EHI). Both implementations were from BoTorch.<sup>65</sup>

The hypervolume improvement identifies the current Pareto front from measured objectives, then quantifies the additional hypervolume contributed by a new candidate relative to a reference point, measuring the improvement across all objectives at each iteration. A detailed formulation is provided by Xu *et al.*<sup>39</sup> For SOBO, where only one objective is optimized, the improvement is computed as the difference between the best achieved value and the reference point, rather than a hypervolume. All reference points were chosen *via* the `infer_reference_point` function from BoTorch.<sup>65</sup>

We chose random selection of candidates from the available search space as a baseline for comparison. This approach ignores prior information and mimics unguided exploration, offering a conservative measure against which BO performance can be evaluated.

To quantitatively compare SOBO and MOBO results across different optimization runs, we introduced a normalized performance metric defined as follows. For each iteration, we computed the difference between GNN performance (*e.g.*, hypervolume improvement) and random selection performance as a function of the number of proposed candidates, normalized by the performance gap between the best possible outcome and random selection. The area under the curve (AUC) of this normalized performance curve provides a single score representing how closely the GNN approaches optimal performance relative to random selection. Positive values indicate that the GNN outperforms random selection, negative values the opposite, and zero corresponds to equivalent performance. This metric was calculated for both hypervolume improvement and Pareto-front coverage to ensure consistency across optimization scenarios.

$$\text{AUC} = \int_1^N \frac{\text{GNN}(i) - \text{random}(i)}{\text{best}(i) - \text{random}(i)} d_i \quad (19)$$

## 3 Results & discussion

### 3.1 Comparison between surrogate model types

We start by comparing the performance of the GNN-based surrogate against the more classical GP surrogate for selectivity. Adsorption data from Li *et al.*<sup>36</sup> were used as objectives. These three objectives are adsorbed CH<sub>4</sub>, adsorbed N<sub>2</sub>, and CH<sub>4</sub> over N<sub>2</sub> selectivity ( $S_{\text{CH}_4/\text{N}_2}$ ). These published data are from Grand Canonical Monte Carlo (GCMC) simulations for a 50% binary mixture of CH<sub>4</sub> and N<sub>2</sub> at 1 bar and 298 K (*i.e.*, not Henry coefficients). Adsorptions are in units of mol per kg. Ten independent SOBO runs were conducted using GNN and BO surrogate models, along with random selection to provide a naive baseline selection process (Fig. 3(a)). Each run started from 25 randomly chosen initial samples (Algorithm 1) and proceeded for 850 iterations. Both GNN and GP surrogates significantly outperformed random selection for maximization of  $S_{\text{CH}_4/\text{N}_2}$ . However, none of the GP models identified the highest selectivity MOF within the allocated budget. From approximately 550

sampled MOFs, the GP surrogate ceased improving, indicated by zero standard deviation in acquisition scores. This stagnation reflects local optima rather than repeated sampling of the same MOF. In contrast, all ten GNNs eventually identified the optimal MOF within the budget, typically evaluating about 1% (*i.e.*, 850 suggested structures) of the 77 482 structures. GNNs models experienced a similar temporary stagnation between 200 and 350 iterations but recovered to find superior candidates.

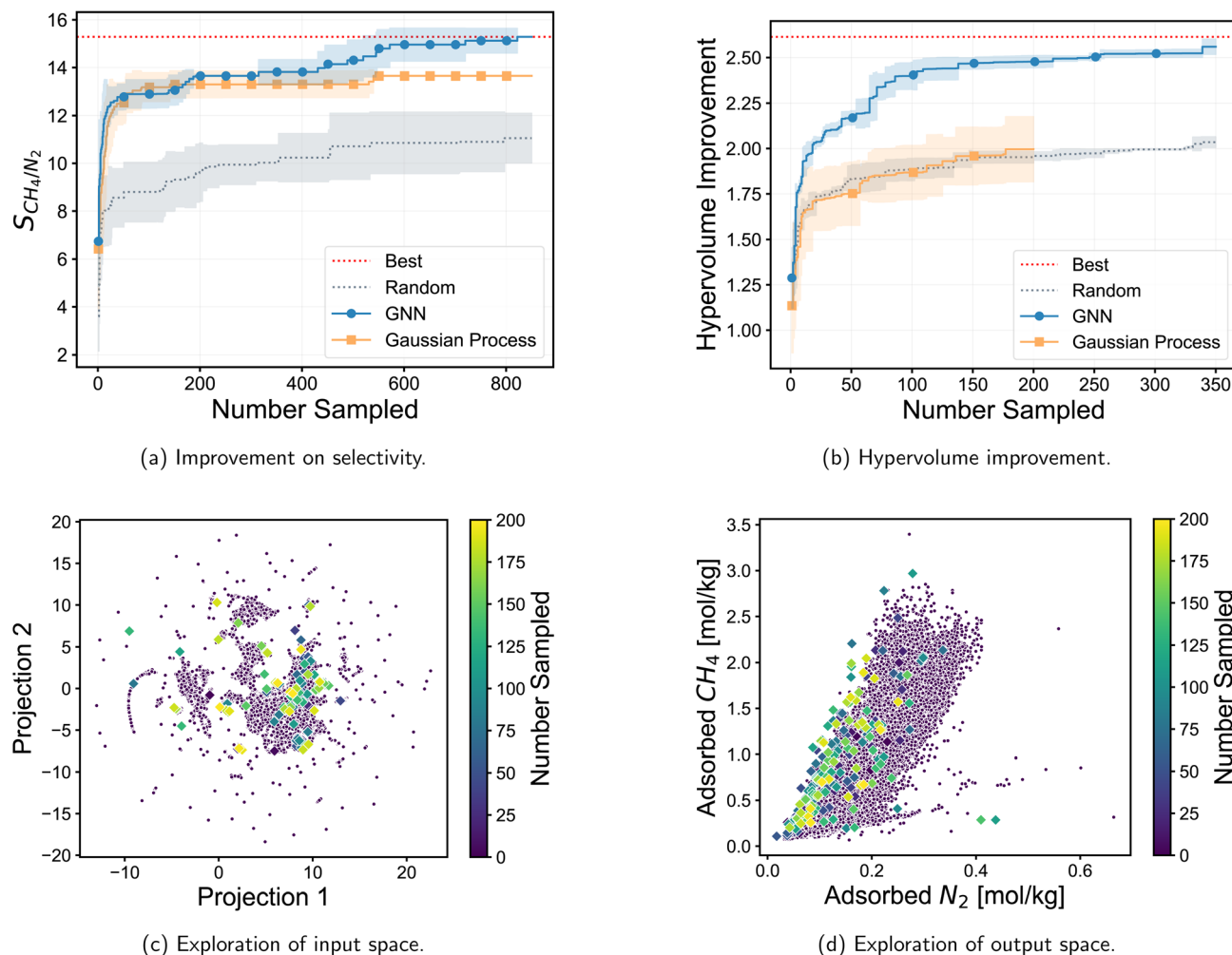
We compared the case for optimizing two objectives (Fig. 3(b)) similar to that of the single objective in Fig. 3(a). Specifically, surrogates were programmed to maximize adsorbed CH<sub>4</sub> and minimize adsorbed N<sub>2</sub>. Three independent runs were averaged instead of ten used for the single objective. It is clear that GNNs outperform GPs and do so faster than the previous comparison. For a randomly chosen GNN surrogate, we show how the model explored both the input and objective space of our MOFs. We only include the first 200 samples to avoid visual clutter. The global level features of MOFs were gathered, transformed using scikit-learn's RobustScaler,<sup>63</sup> and projected onto a two-dimensional (2D) space with the use of Uniform Manifold Approximation and Projection (UMAP)<sup>66</sup> (Fig. 3(c)). The `min_dist` parameter of UMAP was set to 0.01. As seen in Fig. 3(c), there are sparse and dense regions in the low-dimensional 2D UMAP space. The surrogate samples points from both sparse and dense regions, indicating exploration. Fig. 3(d) shows the adsorption objectives on each axis and the same sampling of MOFs from Fig. 3(c). The GNN surrogate has sampled MOFs with objectives in the direction of higher CH<sub>4</sub> adsorption and lower N<sub>2</sub> adsorption (*i.e.*, in the direction of the Pareto front). Identifying MOFs that maximize CH<sub>4</sub> uptake while minimizing N<sub>2</sub> adsorption directly translates to materials that enhance methane recovery and reduce energy costs in pressure-swing adsorption or membrane-based separation processes. Values of 0 on the color maps indicate MOFs that have not been sampled by the GNN. The superior performance of GNN over GP is attributed to its ability to capture complex structure–property relationships beyond the global features used by GPs. Consequently, GNNs surrogates were utilized for all subsequent optimization tasks in this work.

### 3.2 Correlations across Henry coefficients

This subsection focuses on challenging separations of VOCs using MOFs. As established in Section 2.2, maximizing separation involves increasing the Henry coefficient for the favored component while minimizing it for the excluded one. This task becomes difficult when Henry coefficients for both species are highly correlated (*i.e.*, Pearson correlation factors ranging from 0.98 to effectively 1.0 for our VOCs). Although, no single MOF perfectly discriminates between VOCs, the Pareto front identifies the set of optimal trade-offs offering the best achievable selectivities (ratios of Henry coefficients, eqn (2)). We highlight one such Pareto-optimal MOF for later feature sensitivity analysis (Section 3.3).

Strong correlations between Henry coefficients indicate that the adsorption physics are dominated by a scaling relationship





**Fig. 3** Comparison of surrogate models on Bayesian optimization (BO). In panel (a), runs using random selection (dotted gray line), graph-neural network (GNN, blue circles), and Gaussian process (GP, yellow squares) surrogate models were compared for selecting metal–organic frameworks (MOFs) with higher  $CH_4$  over  $N_2$  selectivities. The highest selectivity in the dataset is shown by the horizontal red line and is the optimal material in the search, which is eventually found by all GNN-type but not GP-type surrogate models. Shaded areas are the standard deviation across averaged runs. Panel (b) shows similar information to that of panel (a) with the exception of hypervolume improvement instead of the single objective of selectivity. The dotted red line is the highest possible hypervolume of the dataset. Panel (c) shows the 2D projection of global level descriptors and how a select GNN explored the uniform manifold approximation and projection (UMAP) space, while panel (d) shows the explored set of adsorption values. The color bar for panels (c) and (d) indicates the iteration at which a MOF was sampled by the surrogate. For both of these panels, the diamond symbols indicate MOF structures sampled by the GNN.

where increased binding site density enhances both metrics simultaneously. Successful separation of alkanes from alcohols, for instance, requires minimizing the Henry coefficients of 1-propanol and 2-propanol while maximizing those for butane and propane in the four objective case, representing competing objectives that define a multi-dimensional Pareto front. The Pareto-optimal MOFs identified by our GNN model are those that maximize the ‘residual’ performance, achieving higher selectivity than would be predicted by the adsorption capacity alone. From an industrial perspective, even a small deviation from the dominant trend that would lead to a 2% increase in selectivity can result in significantly lower regeneration costs and higher product purity. Our aim is to develop a surrogate model for BO that can accurately identify the Pareto optimal set and distinguish it from suboptimal solutions.

### 3.3 Performance of surrogate model across Henry coefficients

This subsection validates the predictive accuracy of the GNN surrogate model on the multi-objective optimization Henry coefficient targets. Utilizing 800 randomly sampled MOFs for training, and 100 each for early stopping validation and testing, the model's generalization was assessed. Parity plots for 1-propanol and butane (Fig. 4) demonstrate strong predictive fidelity, with root mean squared error (RMSE) values of 1.355 and 1.491  $\ln([\text{mol kg}^{-1} \text{Pa}^{-1}])$  for 1-propanol and butane, respectively. Considering that the true Henry coefficient values for these VOCs span approximately 11 natural log units, an RMSE of around 1.5  $\ln([\text{mol kg}^{-1} \text{Pa}^{-1}])$  corresponds to an error of approximately 14% relative to the data span, acceptable for



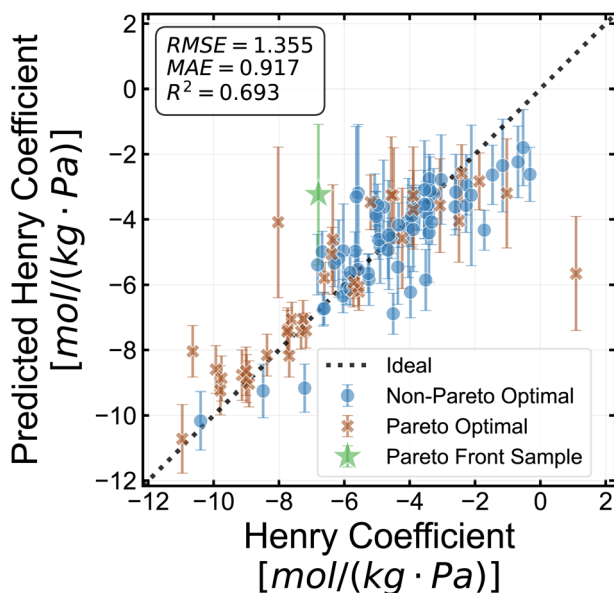
early-stage candidate screening. Similar accuracy for 2-propanol and propane (shown in the SI) confirms robustness.

To evaluate the surrogate's ability to capture meaningful structure–property connections, we measured EHI score for all points. Among the top five scoring candidates, the three highest-scoring were not on the Pareto front, while the remaining two were Pareto-optimal. We selected the highest-scoring of the two Pareto front points (indicated by the star shape in Fig. 4) for the feature importance ablation study. This analysis revealed that node-level features predominantly drive model predictions, with eight of the top ten most important features arising from node descriptors and only two from global features for both 1-propanol and butane. The absence of edge features among the top contributors highlights the dominant

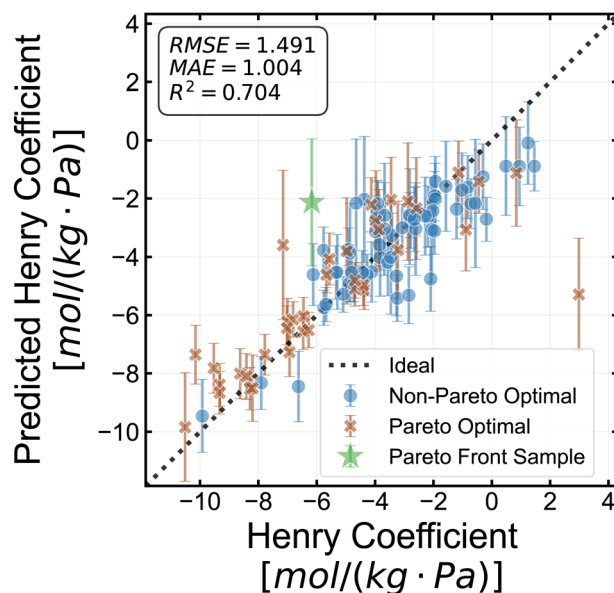
role of local atomic environments captured at the node level. This underscores the limitation of surrogate models restricted to tabular global features, which would inherently miss significant predictive information embedded in the graph topology and atomic features. These results validate the surrogate's reliability and justify its use within the BO framework to effectively navigate strongly correlated multi-objective MOF design spaces.

### 3.4 Comparison on Bayesian optimization and the number of optimized objectives

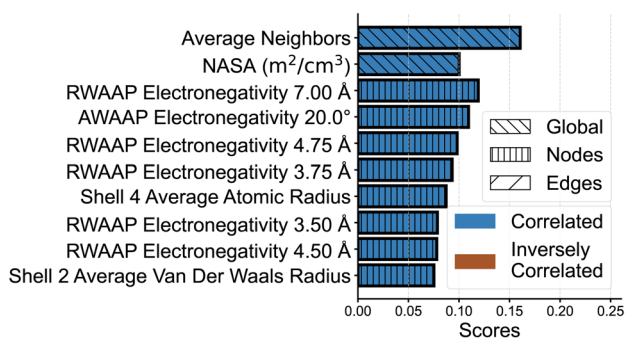
The results presented in the previous sections demonstrated that the GNN surrogate can reliably predict multiple objectives



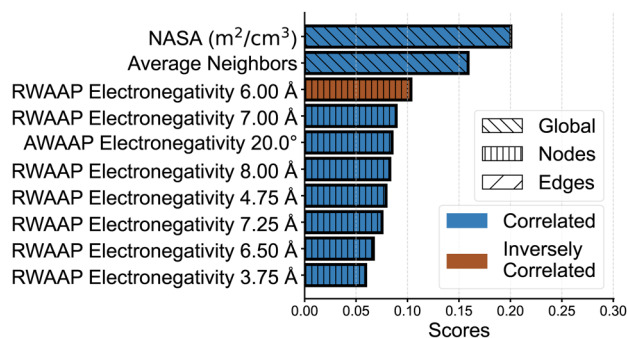
(a) Parity plot for 1-propanol.



(b) Parity plot for butane.



(c) Ablation for 1-propanol.



(d) Ablation for butane.

**Fig. 4** Parity plots and feature ablation analysis. Panels (a) and (b) present parity plots illustrating the performance of a single surrogate model in predicting each objective. The units are in a natural logarithm. The single star green point represents the Pareto front sample used to explain feature significance. The blue circles and brown crosses are the non-Pareto front and Pareto front points, respectively. Panels (c) and (d) display the top 10 features from a feature ablation study for the green star point. Average neighbors, non-available surface area (NASA), radial weighted average atomic property (RWAAP), angular weighted average atomic property (AWAAP), and shell features most influence model predictions. In these plots, blue bars indicate that increasing a feature value during ablation leads to an increase in the model prediction (*i.e.*, correlated), while brown bars indicate that increasing the feature value results in a decrease in the model prediction (*i.e.*, inversely correlated). Bar patterns correspond to the type of graph feature (*e.g.*, global, node, or edge). Scores were normalized such that the sum of all is one.

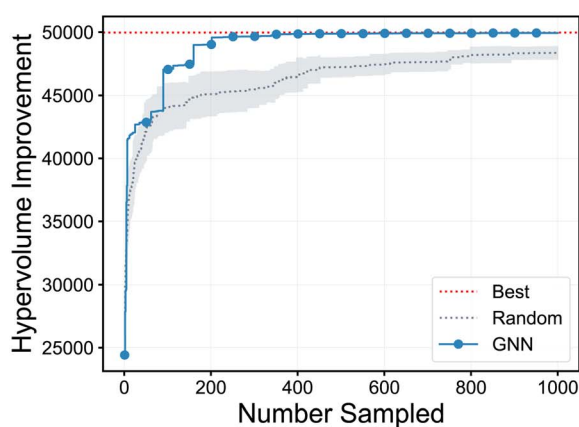


associated with multiple selectivities in MOFs. Building on these results, we now evaluate the efficacy of our GNN as a surrogate model within BO for one- to four-objective optimization tasks, as summarized in Table 1.

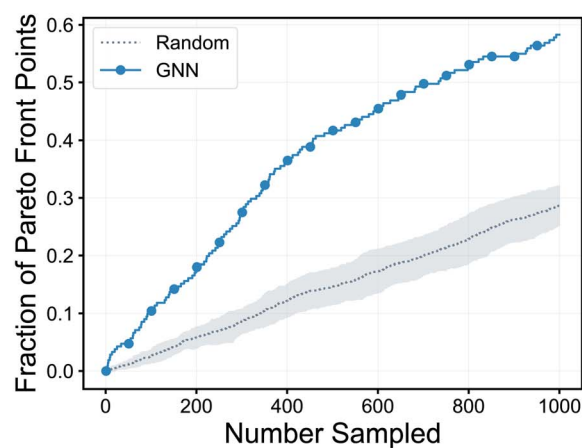
Fig. 5 illustrates results from a representative MOBO run optimizing butane selectivity over 1-propanol and 2-propanol. Panel (a) shows hypervolume improvement *versus* the number of proposed candidates, where the GNN-based surrogate significantly outperforms random selection, achieving near-maximal hypervolume after about 200 candidates, while random selection falls short even after 1000. The random proposals were averaged across 10 runs. Panel (b) shows the fraction of identified Pareto front points instead of hypervolume improvement, with the GNN identifying nearly 60% of

points on the Pareto front after 1000 proposed candidates compared to 26% for random search.

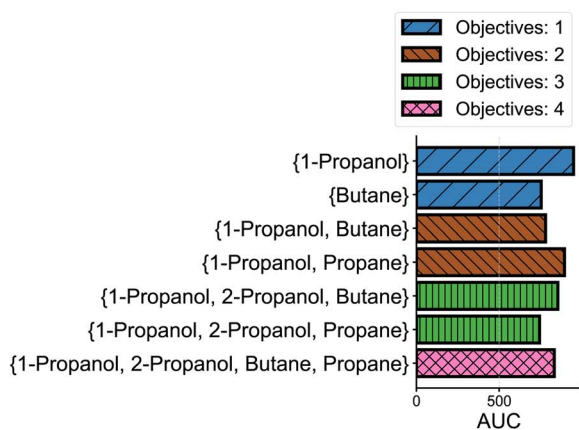
Panels (c) and (d) compare the AUC for normalized hypervolume improvement and Pareto front coverage across runs with one to four objectives. In panel (c), the GNN's AUC of normalized hypervolume improvement, relative to random selection, remains largely consistent across runs optimizing one through four objectives. By contrast, panel (d) shows that the AUC of normalized fraction of Pareto-front points identified is highest for single-objective optimization, similar for two- and three-objective cases, and drops sharply in the four-objective scenario. While normalized hypervolume improvement remains stable across objectives, coverage declines for four-objective optimizations due to the exponential growth of the



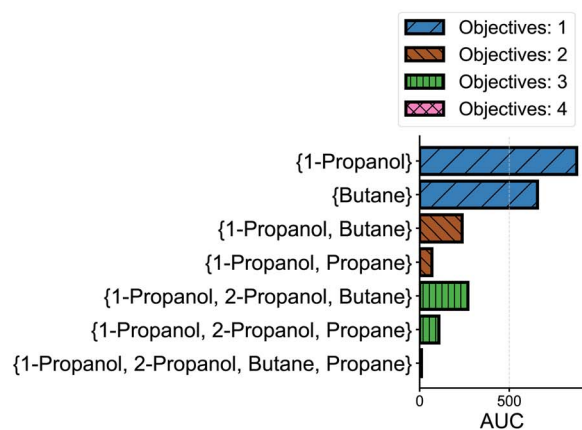
(a) Hypervolume improvement.



(b) Fraction of Pareto front sampled.



(c) AUC for hypervolume improvement.



(d) AUC for fraction of Pareto front sampled.

Fig. 5 Performance evaluation of Bayesian optimization (BO) across varying numbers of objectives. Panel (a) demonstrates hypervolume improvement as a function of proposed candidates. Specifically, the Henry coefficient of butane was maximized while simultaneously minimizing the Henry coefficients for 1-propanol and 2-propanol. The dotted horizontal red line indicates the theoretical maximum hypervolume improvement, the blue line with markers shows the hypervolume improvement achieved through graph-neural network (GNN) selection, and the gray dotted line with a shaded region represents the hypervolume improvement from random acquisition (shaded area denotes standard deviation across 10 runs). Panel (b) displays the fraction of Pareto front points sampled as a function of proposed candidates at each iteration, employing the same color scheme as panel (a). Panel (c) compares the normalized hypervolume improvement across multiple runs with varying numbers of objectives, where overall performance was quantified by computing the area under these curves (larger areas indicate superior optimization performance). Panel (d) presents an analogous analysis for the normalized fraction of Pareto front points identified. The number of objectives is identified by the types of dashes and colors within each horizontal bar.



Table 2 Datasets for Bayesian optimization (BO) with their corresponding number of objectives, metric reported, and area under the curve (AUC)

Dataset	# Objectives	Metric	AUC
{1-Propanol}	1	Hypervolume	942.5
		Fraction	876.5
{Butane}	1	Hypervolume	776.4
		Fraction	681.5
{1-propanol, 2-propanol}	2	Hypervolume	801.3
		Fraction	140.0
{1-propanol, butane}	2	Hypervolume	772.0
		Fraction	236.1
{1-propanol, propane}	2	Hypervolume	893.1
		Fraction	64.8
{2-cis-butene, 2-trans-butene}	2	Hypervolume	824.6
		Fraction	107.4
{Benzene, toluene}	2	Hypervolume	766.8
		Fraction	-7.1
{Butane, propane}	2	Hypervolume	852.5
		Fraction	333.8
{1-propanol, 2-propanol, butane}	3	Hypervolume	867.8
		Fraction	278.7
{1-propanol, 2-propanol, propane}	3	Hypervolume	771.5
		Fraction	112.6
{1-propanol, 2-propanol, butane, propane}	4	Hypervolume	817.4
		Fraction	8.4

Pareto front dimensionality and the intrinsic difficulty of covering it extensively. For example, the single-objective front contains only one optimal point, the three-objective front (shown in panels (a) and (b)) has 211 points, and the four-objective front comprises 1440 points (approximately 40% of the dataset). Despite this, the GNN identified Pareto-front points at rates similar to random selection in higher dimensions, but the specific points it selected yielded substantially greater hypervolume improvement, maintaining superior optimization.

Finally, Table 2 reports GNN BO performance across various VOC separations within the same functional group family, highlighting generally positive AUC values except for the {benzene, toluene} dataset for the metric of fraction, where performance falls to random baseline, likely due to dataset size and the high similarity of these molecules. Overall, the GNN surrogate consistently enables effective optimization for challenging VOC separations in MOFs.

### 3.5 Experimental feasibility of top-ranked MOFs

To bridge the gap between computational prediction and experimental realization, we evaluated the properties of six MOFs and demonstrated how they qualitatively occupy the space of available structural properties. We selected a subset of global graph features commonly used to interpret MOFs and constructed stacked histograms with 15 bins (Fig. 6(a) and (c)) to illustrate the distribution of these properties. Fig. 6(b) presents the top three MOFs on the Pareto front predicted by the GNN model used to generate Fig. 4. These MOFs exhibit large non-accessible surface areas, meaning that adsorbates are unlikely to find suitable uptake sites. Although the GNN surrogate successfully converges to the Pareto front in the four-

objective optimization space, the proposed MOFs show traits that are inconsistent with physically reasonable expectations. Unusual behavior emerges in higher-dimensional spaces where the objectives are strongly correlated and have competing optimization directions. This issue did not arise in the results comparing adsorption values (*i.e.*, not Henry coefficients) from Li *et al.*<sup>36</sup> For the two-objective screening discussed in Section 3.1, the MOFs on the Pareto front selected by the GNN surrogate are shown in Fig. 6(d) and exhibit realistic physical characteristics. For instance, the optimal MOFs possess low non-accessible surface areas and large void spaces (*i.e.*, higher largest free sphere, largest included sphere, and largest included sphere along free path). These trends align with known structure-adsorption relationships, suggesting that the GNN implicitly captures chemically and physically meaningful patterns when identifying optimal materials. The three MOFs in question are DB12-IHUTIQ\_clean\_repeat.cif (Fig. 6(e)), DB0-m3\_o490\_o15\_f0\_fsc.sym.2\_repeat.cif (Fig. 6(f)), and DB1-ZnN4-SiF6-irmof6\_No2\_repeat.cif (Fig. 6(g)). We emphasize that our method screens and finds candidates on the Pareto front, but the Pareto front may lack physically reasonable characteristics which is especially true for higher dimension objective spaces.

Among the three optimal candidates in Fig. 6(d), DB12-IHUTIQ\_clean\_repeat.cif appears to be the most synthetically feasible. This material likely originates from the CCDC database and has been previously reported in the literature.<sup>68</sup> Its organic linker, 1,4-bis(4-pyridyl)-2,3-diaza-1,3-butadiene, is commercially available, overcoming one of the most common limitations in MOF synthesis—linker accessibility. The second candidate, DB0-m3\_o490\_o15\_f0\_fsc.sym.2\_repeat.cif, also appears synthetically feasible. It consists of a bridging tetra-topic carboxylate linker and a pillared N-based linker, both of



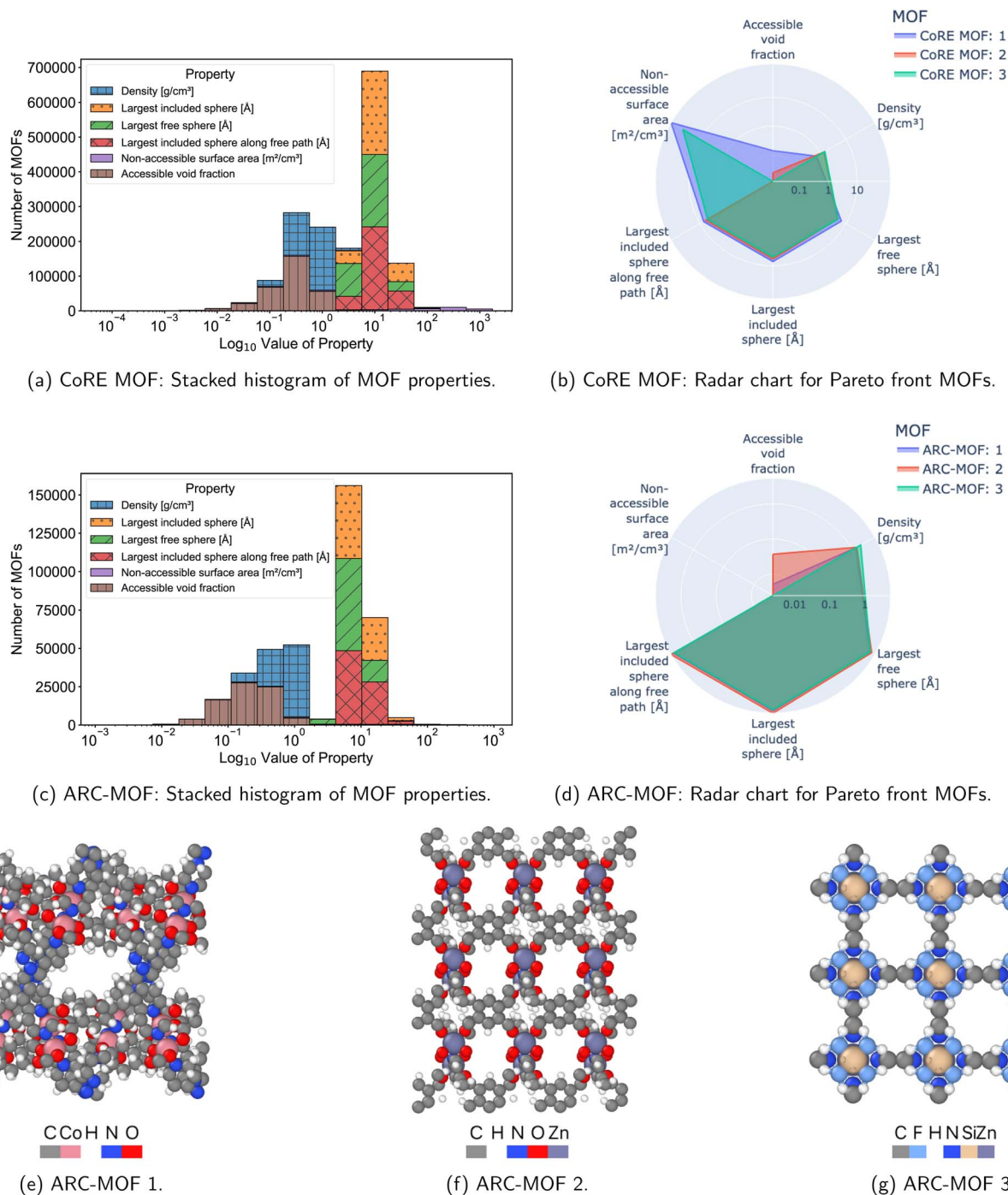


Fig. 6 Selected MOF properties, their span, and values for Pareto optimal metal–organic framework (MOFs). Panel (a) shows a stacked histogram of properties for CoRE MOF shown in the legend. The horizontal axis is the base 10 logarithm of values (with values of zero kept as zero) while the vertical axis represents the counts. Panel (b) shows the property values for MOFs selected by a graph-neural network (GNN). Panels (c) and (d) show the same type of information as panels (a) and (b), respectively, but for the ARC-MOF instead. Panels (e)–(g) show the OVITO<sup>67</sup> visualizations of MOFs from panel (d).

which are available commercially. Although this structure has not been previously reported, its composition suggests it could be readily synthesized. In contrast, DB1-ZnN4-SiF6-irmof6\_No2\_repeat.cif is the least feasible of the three. While conceptually synthesizable, the linker has not been previously synthetically realized. Nevertheless, this structure is isoreticular with the SIFSIX family of materials, implying that MOF

synthesis would be straightforward if the linker were accessible.<sup>69</sup>

## 4 Conclusions

This work demonstrates how detailed structure–property relationships of MOFs can be effectively integrated into a BO



workflow using GNNs. By representing MOFs as graphs with multi-scale features encoded at node, edge, and global levels, GNNs-based surrogates surpass traditional models such as GPs limited to global, tabular representations. Utilizing expected improvement and expected hypervolume improvement acquisition functions for single- and multi-objective optimization respectively, the GNN-enabled optimization efficiently navigates complex design spaces, achieving faster hypervolume gains, notably in challenging separations such as CH<sub>4</sub> over N<sub>2</sub> selectivity, when compared to GP-based approaches.

Despite the inherent strong correlations among Henry coefficients of VOCs studied, GNNs identified Pareto-optimal MOFs with superior performance compared to random search. Ablation analyses underscored the critical role of node-level features in capturing local atomic environments, which are largely overlooked by models restricted to macroscopic descriptors. These results highlight the fundamental advantage of graph-based representations in encoding physicochemically relevant information essential for accurate surrogate modeling and effective candidate selection in materials optimization tasks.

Moving forward, integrating dynamic or flexible MOF representations that capture framework flexibility and adsorption site heterogeneity could improve realism and predictive accuracy. Incorporating multi-fidelity simulation data or experimental feedback into the BO loop may accelerate convergence and robustness of predictions. Enhancing uncertainty quantification methods within GNN architectures will better guide exploration–exploitation trade-offs. Lastly, extending this workflow to optimize additional properties such as adsorption kinetics or stability metrics would broaden its applicability to real-world materials design challenges.

## Author contributions

R. D. and L. E. S. conceived the idea. L. E. S. featurized MOFs as graphs, trained and optimized GNNs, developed the BO routine, and wrote significant portions of this article. N. G. and N. S. B. performed simulations to measure Henry coefficients for VOCs, wrote the corresponding methods, and reviewed this article. D. F. S. G. and R. D. supervised the project and contributed to the draft of this article.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

All raw data, processed data, and computational codes generated in this study are available from the corresponding author, Rémi Dingreville, upon reasonable request at [rdingre@sandia.gov](mailto:rdingre@sandia.gov). The CoRE MOF 2019 database (Version 1.1.3) is publicly available at <https://zenodo.org/record/3677685> and the ARC-MOF (V6) database is accessible at <https://doi.org/10.5281/zenodo.13891643>. Previously published data used to validate

our methods are available at <https://github.com/lwx199906/MOF-GRU.git>.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d5ta09133k>.

## Acknowledgements

This work was supported by the Laboratory Directed Research and Development Program at Sandia National Laboratories. Computational capabilities were supported by the Center for Integrated Nanotechnologies (CINT), an Office of Science User Facility operated for the U.S. Department of Energy (DOE), Office of Science. This article has been authored by an employee of National Technology & Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan <https://www.energy.gov/downloads/doe-public-access-plan>.

## References

- 1 V. F. Yusuf, N. I. Malek and S. K. Kailasa, *ACS Omega*, 2022, 7, 44507–44531.
- 2 H. Furukawa, K. E. Cordova, M. O’Keeffe and O. M. Yaghi, *Science*, 2013, 341, 1230444.
- 3 H.-C. J. Zhou and S. Kitagawa, *Chem. Soc. Rev.*, 2014, 43, 5415–5418.
- 4 G. Li, H. Kobayashi, J. M. Taylor, R. Ikeda, Y. Kubota, K. Kato, M. Takata, T. Yamamoto, S. Toh, S. Matsumura and H. Kitagawa, *Nat. Mater.*, 2014, 13, 802–806.
- 5 J.-R. Li, R. J. Kuppler and H.-C. Zhou, *Chem. Soc. Rev.*, 2009, 38, 1477–1504.
- 6 S. K. Firooz and D. W. Armstrong, *Anal. Chim. Acta*, 2022, 1234, 340208.
- 7 L. E. Kreno, K. Leong, O. K. Farha, M. Allendorf, R. P. Van Duyne and J. T. Hupp, *Chem. Rev.*, 2012, 112, 1105–1125.
- 8 S. Okur, T. Hashem, E. Bogdanova, P. Hodapp, L. Heinke, S. Bräse and C. Wöll, *ACS Sens.*, 2024, 9, 622–630.
- 9 J. Rocha, L. D. Carlos, F. A. A. Paz and D. Ananias, *Chem. Soc. Rev.*, 2011, 40, 926–940.
- 10 J. I. Deneff, L. E. S. Rohwer, K. S. Butler, B. Kaehr, D. J. Vogel, T. S. Luk, R. A. Reyes, A. A. Cruz-Cabrera, J. E. Martin and D. F. Sava Gallis, *Nat. Commun.*, 2023, 14, 981.
- 11 Y. Hu, Z. Huang, L. Zhou, D. Wang and G. Li, *J. Sep. Sci.*, 2014, 37, 1482–1488.
- 12 J. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S. T. Nguyen and J. T. Hupp, *Chem. Soc. Rev.*, 2009, 38, 1450–1459.



- 13 R. E. Sikma, D. J. Vogel, R. A. Reyes, M. L. Meyerson, P. G. Kotula and D. F. Sava Gallis, *Adv. Mater.*, 2024, **36**, 2407435.
- 14 J. Della Rocca, D. Liu and W. Lin, *Acc. Chem. Res.*, 2011, **44**, 957–968.
- 15 F. Ke, Y.-P. Yuan, L.-G. Qiu, Y.-H. Shen, A.-J. Xie, J.-F. Zhu, X.-Y. Tian and L.-D. Zhang, *J. Mater. Chem.*, 2011, **21**, 3843–3848.
- 16 D. F. Sava Gallis, K. S. Butler, J. O. Agola, C. J. Pearce and A. A. McBride, *ACS Appl. Mater. Interfaces*, 2019, **11**, 7782–7791.
- 17 P. Falcaro, R. Ricco, A. Yazdi, I. Imaz, S. Furukawa, D. MasPOCH, R. Ameloot, J. D. Evans and C. J. Doonan, *Coord. Chem. Rev.*, 2016, **307**, 237–254.
- 18 C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat. Chem.*, 2012, **4**, 83–89.
- 19 Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.
- 20 J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. G. Boyd, S. Maley, M. Gibaldi, S. Simrod, V. Ogden and T. K. Woo, *Chem. Mater.*, 2023, **35**, 900–916.
- 21 J. Burner, J. Luo, A. White, A. Mirmiran, O. Kwon, P. G. Boyd, S. Maley, M. Gibaldi, S. Simrod and T. K. Woo, *ab initio REPEAT Charge MOF Database (ARC-MOF)*, 2025, DOI: [10.5281/zenodo.13891643](https://doi.org/10.5281/zenodo.13891643).
- 22 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl and R. Q. Snurr, *J. Chem. Eng. Data*, 2019, **64**, 5985–5998.
- 23 A. White, M. Gibaldi, J. Burner, R. A. Mayo and T. Woo, *Alarming structural error rates in MOF databases used in data driven workflows identified via a novel metal oxidation state-based method*, <https://chemrxiv.org/engage/chemrxiv/article-details/6706b96312ff75c3a1fb0365>.
- 24 Z. Yao, B. Sánchez-Lengeling, N. S. Bobbitt, B. J. Bucior, S. G. H. Kumar, S. P. Collins, T. Burns, T. K. Woo, O. K. Farha, R. Q. Snurr and A. Aspuru-Guzik, *Nat. Mach. Intell.*, 2021, **3**, 76–86.
- 25 Z. Han, Y. Yang, J. Rushlow, J. Huo, Z. Liu, Y.-C. Hsu, R. Yin, M. Wang, R. Liang, K.-Y. Wang and H.-C. Zhou, *Chem. Soc. Rev.*, 2025, **54**, 367–395.
- 26 R. Wang, Y. Zhong, L. Bi, M. Yang and D. Xu, *ACS Appl. Mater. Interfaces*, 2020, **12**, 52797–52807.
- 27 T. D. Pham and R. Q. Snurr, *Langmuir*, 2025, **41**, 4585–4593.
- 28 A. Deshwal, C. M. Simon and J. R. Doppa, *Mol. Syst. Des. Eng.*, 2021, **6**, 1066–1086.
- 29 N. Gantzler, A. Deshwal, J. R. Doppa and C. M. Simon, *Digital Discovery*, 2023, **2**, 1937–1956.
- 30 T.-W. Liu, Q. Nguyen, A. B. Dieng and D. A. Gómez-Gualdrón, *Chem. Sci.*, 2024, **15**, 18903–18919.
- 31 Y. Comlek, T. D. Pham, R. Q. Snurr and W. Chen, *npj Comput. Mater.*, 2023, **9**, 1–14.
- 32 Y. G. Chung, D. A. Gómez-Gualdrón, P. Li, K. T. Leperi, P. Deria, H. Zhang, N. A. Vermeulen, J. F. Stoddart, F. You, J. T. Hupp, O. K. Farha and R. Q. Snurr, *Sci. Adv.*, 2016, **2**, e1600909.
- 33 A. W. Thornton, C. M. Simon, J. Kim, O. Kwon, K. S. Deeg, K. Konstas, S. J. Pas, M. R. Hill, D. A. Winkler, M. Haranczyk and B. Smit, *Chem. Mater.*, 2017, **29**, 2844–2854.
- 34 M. Pardakhti, E. Moharrerri, D. Wanik, S. L. Suib and R. Srivastava, *ACS Comb. Sci.*, 2017, **19**, 640–645.
- 35 J. Luo, O. B. Said, P. Xie, M. Gibaldi, J. Burner, C. Pereira and T. K. Woo, *npj Comput. Mater.*, 2024, **10**, 224.
- 36 W. Li, Y. Situ, L. Ding, Y. Chen and Q. Yang, *ACS Appl. Mater. Interfaces*, 2023, **15**, 59887–59894.
- 37 Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, K. D. Vogiatzis, S. Ling, M. Milisavljevic, H. Zhang, J. S. Camp, B. Slater, J. Ilja Siepmann, D. S. Sholl and R. Q. Snurr, *Computation-Ready Experimental Metal–Organic Framework (CoRE MOF) 2019 Dataset*, <https://zenodo.org/record/3677685>.
- 38 K. S. Walton and D. S. Sholl, *AICHE J.*, 2015, **61**, 2757–2762.
- 39 S. Xu, J. Zhang, H. Chen, Y. Gao, Y. Gao, H. Gao and X. Jia, *Appl. Sci.*, 2022, **13**, 352.
- 40 D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81–101.
- 41 N. S. Bobbitt, R. E. Sikma, J. P. Sammon, M. Chandross, J. I. Deneff and D. F. Sava Gallis, *ACS Sens.*, 2025, **10**, 360–375.
- 42 C. J. Casewit, K. S. Colwell and A. K. Rappe, *J. Am. Chem. Soc.*, 1992, **114**, 10046–10053.
- 43 B. Chen, J. J. Potoff and J. I. Siepmann, *J. Phys. Chem. B*, 2001, **105**, 3093–3104.
- 44 N. Rai and J. I. Siepmann, *J. Phys. Chem. B*, 2007, **111**, 10790–10799.
- 45 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.
- 46 C. D. Wick, M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 2000, **104**, 8008–8016.
- 47 C. E. Wilmer, K. C. Kim and R. Q. Snurr, *J. Phys. Chem. Lett.*, 2012, **3**, 2506–2511.
- 48 N. S. Bobbitt, R. E. Sikma, J. P. Sammon, M. Chandross, J. I. Deneff and D. F. Sava Gallis, *ACS Sens.*, 2025, **10**, 360–375.
- 49 J. Gonzalez, K. Mukherjee and Y. J. Colón, *J. Chem. Eng. Data*, 2023, **68**, 291–302.
- 50 D. Banerjee, C. M. Simon, A. M. Plonka, R. K. Motkuri, J. Liu, X. Chen, B. Smit, J. B. Parise, M. Haranczyk and P. K. Thallapally, *Nat. Commun.*, 2016, **7**, ncomms11831.
- 51 S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson and G. Ceder, *Comput. Mater. Sci.*, 2013, **68**, 314–319.
- 52 O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo and A. Tropsha, *Nat. Commun.*, 2017, **8**, 15679.
- 53 A. P. Bartók, R. Kondor and G. Csányi, *Phys. Rev. B*, 2013, **87**, 184115.
- 54 L. Mentel, *mendeleev – A Python resource for properties of chemical elements, ions and isotopes*, 2014, <https://github.com/lmentel/mendeleev>.
- 55 M. Fernandez, N. R. Trefiak and T. K. Woo, *J. Phys. Chem. C*, 2013, **117**, 14095–14105.
- 56 J. Behler, *J. Chem. Phys.*, 2011, **134**, 074106.



- 57 M. Gastegger, L. Schwiedrzik, M. Bittermann, F. Berzsenyi and P. Marquetand, *J. Chem. Phys.*, 2018, **148**, 241709.
- 58 S. Kancharlapalli, A. Gopalan, M. Haranczyk and R. Q. Snurr, *J. Chem. Theory Comput.*, 2021, **17**, 3052–3064.
- 59 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.
- 60 A. A. Hagberg, D. A. Schult and P. J. Swart, *Proceedings of the 7th Python in Science Conference*, Pasadena, CA USA, 2008, pp. 11–15.
- 61 M. Fey and J. E. Lenssen, *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
- 62 P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li and R. Pascanu, Relational inductive biases, deep learning, and graph networks, *arXiv*, 2018, preprint, arXiv:1806.01261, DOI: [10.48550/arXiv.1806.01261](https://doi.org/10.48550/arXiv.1806.01261), <http://arxiv.org/abs/1806.01261>.
- 63 F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- 64 T. Akiba, S. Sano, T. Yanase, T. Ohta and M. Koyama, *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.
- 65 M. Balandat, B. Karrer, D. R. Jiang, S. Daulton, B. Letham, A. G. Wilson and E. Bakshy, *Adv. Neural Inf. Process. Syst.*, 2020, **33**, 21524–21538.
- 66 L. McInnes, J. Healy, N. Saul and L. Grossberger, *J. Open Source Softw.*, 2018, **3**, 861.
- 67 A. Stukowski, *Model. Simulat. Mater. Sci. Eng.*, 2009, **18**, 015012.
- 68 J. Ye, X. Li, Z. liang Xu and H. Xu, *J. Mol. Struct.*, 2015, **1093**, 162–165.
- 69 O. Shekhah, Y. Belmabkhout, Z. Chen, V. Guillermin, A. Cairns, K. Adil and M. Eddaoudi, *Nat. Commun.*, 2014, **5**, 4228.

