A process-level perspective of the impact of molecular force fields on the computational screening of MOFs for carbon capture

Conor Cleeton\textsuperscript{1*}, Felipe Lopes Oliveira\textsuperscript{2,3}, Rodrigo F. Neumann\textsuperscript{2}, Amir H. Farmahini\textsuperscript{1}, Binquan Luan\textsuperscript{5}, Mathias Steiner\textsuperscript{2}, Lev Sarkisov\textsuperscript{1}

\textsuperscript{1} The University of Manchester, Manchester, United Kingdom
\textsuperscript{2} IBM Research, Av. República do Chile, 330, CEP 20031-170, Rio de Janeiro, RJ, Brazil
\textsuperscript{3} Instituto de Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil
\textsuperscript{5} IBM Research, 1101 Kitchawan Road, Yorktown Heights, 10519, NY, USA

E-mail: conor.cleeton@postgrad.manchester.ac.uk

*Correspondence concerning this article should be addressed to Conor Cleeton
Abstract

The question we pose in this study is to what extent the ranking of metal organic frameworks (MOFs) for adsorption-based carbon capture, and the selection of top performers identified in PSA process modelling, depends on the choice of the commonly available forcefields. To answer this question, we first generated distributions of CO$_2$ and N$_2$ adsorption isotherms via molecular simulation in 690 MOFs using six typical forcefields: the UFF or Dreiding sets of Lennard-Jones parameters, in combination with partial charges derived from *ab initio* calculations (DDEC scheme) or by charge equilibration (Qeq an EQeq schemes). We then conducted a systematic uncertainty quantification study using PSA process-level modelling. We observe that: (i) the ranking of MOFs significantly depends on the choice of forcefield; (ii) partial charge assignment is the prevailing source of uncertainty, and that charge equilibration schemes produce results which are inconsistent with *ab initio*-derived charges; iii) the choice of Lennard-Jones parameters is a considerable source of uncertainty. Our work highlights that is not really possible to obtain material rankings with high resolution using a single molecular modelling approach and that, as a minimum, some uncertainty should be estimated for the performance of MOFs shortlisted using high throughput computational screening workflows. Future prospects for computational screening studies are also discussed.

**Keywords:** materials screening, metal organic framework, pressure swing adsorption, carbon capture, molecular simulation
1 INTRODUCTION

High throughput computational screening (HTCS) has emerged as an important strategy in the efforts to identify the most promising metal organic frameworks (MOFs) for carbon capture applications. The underlying assumption is that the results obtained using these workflows can be used to inform experimentalists on which MOFs to synthesise and test. Consider two academic groups conducting computational screening to determine the best MOF for carbon capture from flue gas, as an example. Their simulation protocols are identical, including the database of MOF candidates. To rank the performance of materials, both groups use process modelling, with an identical process configuration and parameters. The only difference between the groups is the choice of forcefields used to describe intermolecular interactions, which are required by molecular simulations to generate equilibrium adsorption data. In this article, we question whether the rankings produced by the two groups will be similar and to what extent.

Let us first look in more detail at the steps and the assumptions typically involved in HTCS workflows. Most HTCS studies rely on Grand Canonical Monte Carlo (GCMC) simulations to generate adsorption data. In the multiscale HTCS approach, this data can be used as input to detailed models of pressure swing adsorption (PSA) cycles to evaluate the potential of an adsorbent in a setting which closely resembles the real industrial separation. Key performance indicators (KPIs) at the process level are typically optimised for a given material by varying the combinations of PSA cycle variables (pressures of each step, duration of each step, etc.). We are often interested in multiple competing KPIs, such as the CO₂ product purity and recovery, or the energy penalty (energy required to capture a unit amount of CO₂) and productivity (how much CO₂ is captured per kg of adsorbent per second). This trade-off between competing objectives can be represented by a series of Pareto-optimal solutions known as a Pareto front, and comparing Pareto fronts for two different materials is an established protocol for ranking materials based on their performance. In the process of constructing a simulation pipeline that can predict the
performance of a MOF simply from its crystal structure, one is posed with several important design choices. The first choice is how to generate the adsorption data using molecular simulations. To understand the importance of this step, it is instructive to briefly review the primary components of a typical GCMC simulation of adsorption in MOFs, and the possible options available to practitioners in the field.

In order to predict the adsorption of CO\textsubscript{2} and N\textsubscript{2} in a MOF, the potential energy of the system, and therefore the interactions between all atoms, must be accurately described. In classical GCMC, this is typically achieved by defining an appropriate set of equations and molecular parameters known as a *forcefield* (FF), whereby short-range repulsion / dispersion interactions are described by a Lennard-Jones (LJ) potential, and long-range electrostatic interactions by a Coulombic potential.\textsuperscript{9,10} Specialised FFs have been developed to accurately describe the LJ contributions in specific subclasses of MOFs,\textsuperscript{9,11,12} however these lack the large-scale predictive capabilities needed for screening purposes. On the other hand, generic FFs such as Dreiding\textsuperscript{13} and UFF\textsuperscript{14} are transferable (albeit less accurate) as they describe the same types of framework atoms in many different materials with identical potential parameters. Therefore, Dreiding or UFF are the common and practical choices of LJ parameters for HTCS of MOFs.\textsuperscript{15–19} They are typically combined with the TraPPE\textsuperscript{20} FF (which describes the CO\textsubscript{2} and N\textsubscript{2} adsorbates) using mixing rules such as Lorentz-Berthelot. To model the electrostatic interactions in GCMC simulations, partial charges must be assigned to each atom in the system. Several assignment schemes have been developed so far and have been reviewed elsewhere,\textsuperscript{21} each differing in their level of accuracy, philosophy, and computational cost. The most accurate (but computationally expensive) approach would be to determine the electrostatic potential within a material using quantum mechanical (QM) calculations and then assign point charges that best represents this electronic environment.\textsuperscript{21} Among the many QM methods that exist,\textsuperscript{22–30} density derived electrostatic and chemical (DDEC)\textsuperscript{25–29} charges and REPEAT charges\textsuperscript{30} are considered to be the best options for periodic structures. Less accurate, but computationally inexpensive,
charges can also be obtained using charge equilibration schemes such as Qeq,\textsuperscript{31} EEq\textsuperscript{32} and PQeq,\textsuperscript{33} among others.\textsuperscript{34,35} These approaches are based on the principle of electronegativity equalisation between bonded atoms, and have been used extensively in MOF screening studies despite their semiempirical nature.\textsuperscript{15,36–39}

It is currently nontrivial to obtain a set of LJ parameters capable of screening thousands of MOFs with high accuracy. There is also no unified operator that can assign partial charges that reproduce the actual electron density of porous solids, given they are not experimentally observable properties.\textsuperscript{21} It is therefore quite common to see an almost arbitrary combination of the LJ parameters and charge assignment schemes used in HTCS of MOFs, based on the parameter availability and other practical factors. Naturally, this raises some concerns on the quality of the FFs used for predicting equilibrium adsorption data in molecular simulations.\textsuperscript{40–42} Several studies have highlighted the substantial quantitative differences that arise between adsorbate uptakes predicted using different charge schemes\textsuperscript{24,36,43,44} and LJ FFs.\textsuperscript{41,45} Others have also shown that the separation performance of PSA-based CO\textsubscript{2} capture processes strongly depends on the competitive adsorption between CO\textsubscript{2} and N\textsubscript{2}.\textsuperscript{46–49} Altogether, this may lead to significant differences in the process-level ranking of porous materials for carbon capture depending on the particular combination of interaction parameters employed. This in turn casts a shadow of doubt on the recommendations produced by theoretical studies, hindering the transition of the recommended materials to the experimental campaign and ultimately to practical, industrial systems.

To explore these issues, we carry out a multiscale HTCS where we assess the suitability of six FFs commonly chosen to simulate CO\textsubscript{2} and N\textsubscript{2} adsorption in MOFs. Our goal is to understand how the material rankings are impacted by this choice, not necessarily on how well the FFs reproduce experimental data, and so we deliberately avoid making any comparisons to experimental measurements of adsorption. As a process-level case study, we consider the separation of CO\textsubscript{2} (15 vol\%) and N\textsubscript{2} (85 vol\%) binary mixtures using a
modified Skarstrom PSA cycle at conditions representative of dry flue gas streams from coal-fired power plants. We generate distributions of CO\(_2\) and N\(_2\) adsorption isotherms in 690 MOFs at multiple temperatures using LJ parameters from either Dreiding or UFF, in combination with partial charges assigned by the DDEC, EEq, or Qeq schemes, as described in sections 2.1 and 2.2. We then determine the distributions in process-level performances for every MOF using a combination of high-fidelity PSA process modelling and machine learning (sections 2.3 and 2.4). Using these data, we generate material ranking lists for each FF and reflect on the consistency of the results. Our analysis will show only moderate-to-poor agreement between the process-level rankings obtained using different FFs, and that this disagreement stems primarily from the significant deviations which manifest at the material level of description. We then quantify the average uncertainties between FFs from their process-level responses (described in section 2.5) and reveal that the prevailing source of uncertainty is the choice of charge scheme. We finally explore a potential pathway towards uncertainty mitigation in multiscale simulations of PSA-based CO\(_2\) capture processes and discuss future prospects for HTCS of materials.

2 METHODOLOGY

2.1 Grand Canonical Monte Carlo simulations

The CoRE MOF 2014 DDEC database\(^{50,51}\) is chosen as the starting point for adsorbent selection. To ensure a fair comparison in GCMC modelling approaches, only MOF structures that contain all the atom types present in both the UFF and Dreiding force fields are included. This filtering process results in a reduced dataset of 726 structures from the original 2932. A further 36 MOFs are removed due to the presence of unbound water molecules, counter-ions and / or hydrogen atoms with incorrect bond lengths / angles, leaving 690 computation-ready structures available for this study.

Simulations of CO\(_2\) and N\(_2\) adsorption are performed using the atomistic Grand Canonical Monte Carlo (GCMC) method implemented in RASPA.\(^{52}\) Interaction energies
between non-bonded atoms are modelled using the Lennard-Jones (LJ) plus Coulomb potential (Eq. (1)):

\[ U_{ij}(r_{ij}) = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}} \]

\[ \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \]  

(1)  

(2)

Where \( i \) and \( j \) are indices of the interacting atoms, \( r_{ij} \) is the distance between atoms \( i \) and \( j \), \( \varepsilon \) and \( \sigma \) are the LJ parameters, \( \varepsilon_0 \) is the vacuum permittivity, and \( q \) is the partial charge. For framework atoms, LJ parameters were taken from UFF\textsuperscript{14} or Dreiding\textsuperscript{13} For CO\textsubscript{2} and N\textsubscript{2} adsorbates, the Transferable Potentials for Phase Equilibria (TraPPE) FF is used.\textsuperscript{20} LJ parameters between unlike atoms are calculated using the Lorentz-Berthelot mixing rules (Eq. (2)), and LJ potentials beyond 12.8 Å are shifted to zero at the cut-off. For the Coulombic interactions, three partial charge assignment schemes are considered here: DDEC,\textsuperscript{25} EQeq,\textsuperscript{32} and Qeq.\textsuperscript{31} DDEC charges are taken from the original CoRE MOF 2014 DDEC database, while the EQeq and Qeq partial charges are calculated using the EQeq v1.1.0 \textsuperscript{53} and RASPA v2.0.47 simulation codes, respectively. The Ewald summation technique was used to calculate the long-range electrostatic interactions with a relative precision of 10\textsuperscript{-6}. All simulations are performed with framework atoms fixed at their crystallographic positions. A simulation supercell with all three dimensions greater than twice the cut-off distance of 12.8 Å is used to satisfy the minimum image convention.

For each adsorbent considered in this work, single-component isotherms are generated for CO\textsubscript{2} and N\textsubscript{2} at 273K, 298K, and 323K on a logarithmic pressure scale between 10\textsuperscript{-3} – 10 bar for CO\textsubscript{2} and 10\textsuperscript{-3} – 1 bar for N\textsubscript{2}. A single simulation run is used for each isotherm to avoid the use of initialisation cycles; starting from the lowest pressure, the final configuration of the preceding pressure point is used as the initial configuration of the next pressure point. 10,000 Monte Carlo cycles are simulated at each pressure point using swap (insertion or deletion with equal probabilities), translation, rotation, and reinsertion moves with probabilities of 0.5, 0.3, 0.1, and 0.1, respectively. Then, adsorbate uptakes are
calculated by first discarding the transient regime of each simulation using the Marginal Standard Error Rule (MSER) and then taking an ensemble average over the equilibrated regime. The GCMC method requires the fugacity of the adsorbing species (or equivalently, the chemical potential). The Peng-Robinson equation of state is utilised to convert the pressure points on the adsorption isotherm into corresponding fugacity values. This approach has been developed and validated elsewhere, while additional GCMC simulation details can be found in the supporting information (SI), supplementary note 1. However, we would like to point out that the gas species of interest can be considered as nearly ideal gases under the conditions of this study.

Using the protocol described above, we generate 18 isotherms for both CO$_2$ and N$_2$: 3 isotherms for each of the 6 possible FF combinations, with a resolution of 11 pressure points per CO$_2$ isotherm and 10 pressure points per N$_2$ isotherm. This generates 36 isotherms per material using 378 unique adsorption points. In total, for the 690 MOFs considered in this work, 24,840 isotherms are generated which span different pressures, temperatures, charge assignment methods, and generic forcefield parameters. We refer to the data generated using this subset of MOFs as the **Charge-dependent, Reproducible, Accessible, Forcefield-dependent, and Temperature-dependent Exploratory Database** of isotherms, or CRAFTED isotherms for brevity. To provide an initial seed for future benchmarking efforts, we make the CRAFTED database (v1.1.1) and scripts available to the broader scientific community in a dedicated Zenodo repository at [https://zenodo.org/record/7689919](https://zenodo.org/record/7689919).

### 2.2 Equilibrium adsorption model

Process-level simulations require an analytical model to determine the equilibrium molar loadings of CO$_2$ / N$_2$ mixtures over various pressure and temperature domains. The dual-site Langmuir (DSL) model (Eq. (3)) is well suited for this purpose. Within the DSL approach, competitive adsorption may be calculated by a simple extension of the single-component DSL models and many practical systems of interest are known to be described by this behaviour.
\[ q_i^{DSL} = \sum_{j} q_{s,j}^i \frac{b_j^i p_i}{1 + \sum_i b_j^i p_i} \]  

(3) \[ b_j^i = b_{0,j}^i \exp \left( -\frac{\Delta H_{i,j}^{ads}}{RT} \right) \]  

(4)

Here, \( i \) refers to adsorbate (either \( \text{CO}_2 \) or \( \text{N}_2 \)), \( j \) refers to adsorption sites 1 or 2, \( q_{s,j}^i \) \([\text{mmol g}^{-1}]\) is the saturation capacity, \( p_i \) \([\text{bar}]\) is the pressure of component \( i \), \( b_{0,j}^i \) \([\text{bar}^{-1}]\) is the pre-exponential factor, and \( \Delta H_{i,j}^{ads} \) \([\text{J mol}^{-1}]\) is the heat of adsorption. The DSL model parameters are determined by nonlinear least-squares regression against the GCMC adsorption data at all three temperatures using the `lmfit.minimize()` function in python.\(^{59}\) We apply the following constraints during the fitting to obtain physically meaningful DSL model parameters. We enforce that the saturation capacity of each site is equal for both \( \text{CO}_2 \) and \( \text{N}_2 \) to satisfy the thermodynamic consistency requirements.\(^{60}\) We designate the \( \text{CO}_2 \) adsorption site 1 as being the stronger adsorbing site by imposing the \( b_1^{\text{CO}_2} > b_2^{\text{CO}_2} \) condition. This constraint reflects the energetic heterogeneity of \( \text{CO}_2 \) adsorption. For \( \text{N}_2 \), in cases where the adsorption isotherm is linear over the pressure range, we adopt the Equal Energy Site (EES) formulation of the extended DSL model.\(^{61,62}\) This formulation recognises the energetic homogeneity of \( \text{N}_2 \) adsorption by setting the interaction energy parameters (i.e., \( b_{0,j}^{\text{N}_2} \) and \( \Delta H_{N_2,j}^{ads} \)) to the same value at both sites.\(^{49,63}\) In cases where \( \text{N}_2 \) adsorption deviates substantially from linearity, this constraint is relaxed, and the fitting protocol used to model \( \text{CO}_2 \) adsorption is utilised. This approach has been validated using binary mixture GCMC simulations (see supplementary note 1.2).

2.3 PSA simulation and optimisation strategy

We consider a modified Skarstrom PSA cycle which consists of five steps: (1) Pressurisation; (2) Adsorption; (3) Heavy reflux; (4) Counter-current depressurisation; and (5) Light reflux, as shown in Figure 1. The cycle variables consist of high (adsorption) and low (evacuation) pressures \((p_H \text{ and } p_L \text{ [bar]})\), durations of the pressurisation, adsorption,
and counter-current depressurisation steps \((t_{\text{pres}}, t_{\text{ads}}, \text{and } t_{\text{CnCDepres}})\), velocity of the feed \((v_{\text{feed}} \text{ [ms}^{-1}\text{]})\), and light and heavy reflux ratios \((\alpha_{\text{LR}} \text{ and } \alpha_{\text{HR}})\). Other important properties highlighted in Figure 1 are the porosity of the bed and pellet \((\varepsilon_B \text{ and } \varepsilon_p)\), and the pellet radius \((r_p \text{ [m]})\). The role and purpose of individual steps are described in detail elsewhere.\(^{64}\) The main feature of this cycle is that it utilises a dual reflux configuration, which avoids the need to pull deep vacuum below the practical limits of industrial vacuum pumps.\(^{65}\) This presents an important advantage over the more commonly used 4-step VSA cycle with light product pressurisation, which often requires unphysically low evacuation pressures to achieve high CO\(_2\) purities and recoveries.\(^{49}\)

To accurately simulate the separation of CO\(_2\) / N\(_2\), a description of the mass, energy, and momentum transfer phenomena taking place in the adsorption column are required. These form a system of partial differential algebraic equations (PDAEs). We follow the approach detailed in our previous work\(^{63}\) which relies on the simulation code of Yancy-Caballero et al.\(^{64}\) to solve this system of equations: the PDAEs are reduced to ordinary differential equations (ODEs) using a spatial discretisation scheme and the resulting set of ODEs are solved using MATLAB’s in-built ode15s solver.\(^{66}\) A “unibed” simulation strategy is adopted,\(^{67}\) whereby the solutions of partial pressures, temperature, and compositions from
one step – obtained through the imposition of appropriate initial and boundary conditions – are used as the initial conditions for the next step in the cycle. The column is initialised at evacuation pressure $p_L$ and 298K, after which the cycle steps are simulated iteratively until Cyclic Steady State (CSS) is achieved. A complete description of the governing equations, model parameters, boundary conditions and mathematical criterion used to determine CSS are provided in supplementary note 2, while other technical details of the PSA simulation are provided elsewhere. We summarise here the key assumptions adopted in our process model as the following: (1) the Ideal Gas law describes the gas phase; (2) the gas phase and solid phase are in thermal equilibrium; (3) mass transfer is controlled by molecular diffusion in the macropores, which is valid for most microporous adsorbents with pore openings greater than 4Å. Furthermore, the mass transfer rate is described by the linear driving force (LDF) approximation; (4) an axially dispersed plug flow model is employed, meaning there are no radial distributions in concentration, pressure, and temperature for both the solid and gas phase; (5) the column is operated adiabatically; (6) the Ergun equation describes the axial pressure drop. These assumptions are largely consistent with the assumptions of several previous studies.

Four key performance indicators (KPIs) are extracted from the CSS solutions, namely: CO$_2$ purity [-], CO$_2$ recovery [-], energy penalty $[kWh \; ton_{CO_2}^{-1}]$, and productivity $[mol_{CO_2} \; kg_{ads}^{-1} \; s^{-1}]$. Regulatory bodies such as the US Department of Energy have specified 95% purity and 90% recovery targets for CO$_2$ capture processes. Viable adsorbents must therefore be able to achieve this separation potential in any given PSA cycle, and so the purity and recovery of CO$_2$ are important metrics to quantify. For the purposes of our investigation, we relax the purity requirement to 90%, and simply refer to the 90% purity and recovery targets as carbon capture and sequestration (CCS) constraints. The energy penalty corresponds to the amount of energy required to capture a unit of CO$_2$, whereas productivity specifies how much CO$_2$ a unit of the adsorption bed (e.g., a kg or m$^3$) can capture per unit of time. These KPIs are related to the operating and capital costs of the
separation, and have been widely adopted in previous screening studies as the indicative metrics of material performance.\(^4\)–\(^8\),\(^71\) Mathematical definitions for each KPI are provided in supplementary note 2.

The performance of the PSA cycle for a given adsorbent is optimised by coupling the PSA process model with the Non-dominated Sorting Genetic Algorithm (NSGA-II).\(^72\)–\(^74\) 11 decision variables ( DVs ) are initially considered for the process optimisation. These included all of the PSA cycle variables as well as \(\varepsilon_B\), \(\varepsilon_p\) and \(r_p\). Parameters \(\varepsilon_B\), \(\varepsilon_p\) and \(r_p\) are included as DVs in light of the recent work by Farmahini et al.\(^75\) who demonstrated that significant performance gains can be obtained by optimising properties of the pellet as well as the cycle variables. We then reduced the dimensionality of the optimisation landscape from 11 DVs to 7 DVs by fixing the values of \(p_L\), \(t_{\text{pres}}\), \(t_{\text{CnCDepres}}\), and \(\alpha_{HR}\) after some preliminary sensitivity analysis. This decision is motivated by the fact that these 4 parameters contributed very little to the overall performance of the cycle; they either converged to the boundaries defined by physical limitations, or they varied within a narrow window of the optimisable range (see supplementary note 3). A summary of the final DVs considered in this work and their operating ranges are provided in Table 1.

<table>
<thead>
<tr>
<th>Table 1. Lower and upper bounds for decision variables used in the optimisation of the modified Skarstrom cycle.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(p_H) ([\text{bar}])</td>
</tr>
<tr>
<td>Lower Bound</td>
</tr>
<tr>
<td>Upper Bound</td>
</tr>
</tbody>
</table>

* Competitive adsorption of CO\(_2\) / N\(_2\) mixtures are well described by the extended DSL model up to 5 bar, see Figure S1.

We conduct two separate optimisation stages in this study. The first is an unconstrained maximisation of the CO\(_2\) purity and recovery. For this stage, the genetic algorithm is terminated after 70 generations. The second stage is a multi-objective optimisation, whereby we maximise the productivity while simultaneously minimising the energy penalty subject to the two CCS constraints: CO\(_2\) purity \(\geq 0.9\) and CO\(_2\) recovery \(\geq \).
0.9. For this stage, the genetic algorithm is terminated after 250 generations. We confirm that 70 and 250 generations for the CO$_2$ purity-recovery and energy-productivity optimisations, respectively, are sufficient to guarantee convergence of the Pareto fronts.

2.4 Machine Learning surrogate model for the prediction of PSA performance indicators.

We developed a machine learning (ML) surrogate model of the modified Skarstrom cycle which can predict the process KPIs for any arbitrary combination of DVs and material properties. We use this surrogate model to expedite the calculation of CO$_2$ purity-recovery and energy penalty-productivity Pareto fronts. Similar to, and guided by, our previous efforts in this domain, we use an artificial neural network (ANN) as our choice of ML model. There are overall 7 DVs and 14 material-specific parameters (crystal density $\rho_s \,[kg\,m^{-3}]$, crystal specific heat capacity $c_{p,s} \,[J\,kg^{-1}K^{-1}]$, and the 12 DSL parameters which describe CO$_2$ and N$_2$ adsorption), resulting in 21 inputs, while the outputs of interest are the 4 KPIs (i.e., CO$_2$ purity, CO$_2$ recovery, energy penalty, and productivity). Appropriate training limits are defined for each of the material-specific parameters to sufficiently cover the distributions of

![Figure 2. The ANN model prediction accuracy compared to the high-fidelity PSA process model. Left subplots: Parity plots showing the ANN model predictions of individual cycle configurations for each process KPI, along with the corresponding $R^2_{adj}$ values. Each ANN model is trained on 300,000+ datapoints. Right subplot: Examples of energy-productivity Pareto front predictions using the ANN model (solid lines) compared to the Pareto fronts obtained from process simulations (filled circles) for different hypothetical adsorbent materials. Hypothetical adsorbent materials are characterised by random combinations of the material-specific parameters. Shaded region indicates an error of ± 5%.](image-url)
CRAFTED MOF material properties, and training limits for the 7 DVs are defined as the optimisable ranges given in Table 1.

Training, testing, and validation data is generated by a combination of Latin Hypercube sampling (LHS) of the input phase space as well as a guided search of high quality, Pareto-optimal points through a bootstrap optimisation technique. For each combination of inputs identified using these sampling techniques, the high-fidelity PSA process model is simulated to determine the 4 KPIs. Then, the data is cleaned and pre-processed by applying log-like transformations to each of the KPIs which improved the overall accuracy of the ANN model. Once the data is prepared, it is partitioned into train / test / validation datasets using a ratio of 90 / 5 / 5. MATLAB’s machine-learning toolbox is subsequently used to train a dense feedforward ANN model architecture of 3 hidden layers with 45 neurons per hidden layer using the Levenberg-Marquardt back-propagation. Once trained, the ANN model predictive performance is evaluated by calculating the adjusted coefficient of determination ($R^2_{adj}$) for all KPIs using the test dataset. Model training and refinement is terminated once predictions of both the individual cycle configurations and fully resolved Pareto fronts no longer improved (Figure 2). Additional details on the surrogate model training protocol and on how the CRAFTED MOF material properties are calculated are provided in the supplementary note 4.

2.5 Uncertainty metrics for high performing adsorbents

High performing adsorbents are characterised by their ability to minimise the cost of CO$_2$ capture while meeting the CCS requirements. It is not a priori evident what combination of material properties leads to these requirements being satisfied, so we conduct preliminary purity-recovery optimisations in two stages. Low performing adsorbents are first identified by their ANN-generated Pareto fronts and discarded from our subsequent analysis. We define a low performing material as one which does not satisfy the CCS constraints for any FF. For the 218 materials that remain after this pre-screening, high quality Pareto data is generated by refining the ANN results with the high-fidelity process model for an additional 10 NSGA-II
generations. This mitigates any potential bias in the ML process and ensures that the conclusions drawn are established using accurate process simulation data. Moving forwards, we only consider the 218 MOFs in our uncertainty calculations. We designate this subset of MOFs as CRAFTED-u.

We seek to quantify the level of agreement between FFs at the molecular and process simulation levels. To do so, we make use of a number of statistical metrics which are both local, i.e., relating to a single material, and global, i.e., relating to the entire CRAFTED-u MOF database. Let us first focus on molecular simulations, used to generate equilibrium adsorption data. In addition to the comparison of individual adsorption isotherms generated from different FFs, we also employ the Spearman correlation coefficient, \( \rho \), to measure the global extent of the correlation between adsorbate uptakes predicted by different FFs. Perfect positive correlations result in \( \rho = 1 \), perfect negative correlations in \( \rho = -1 \), and no correlation in \( \rho = 0 \).

**Figure 3.** High throughput computational screening workflow with uncertainty quantification. For a single CRAFTED MOF, comparing two different molecular forcefields \( i \) and \( j \) involves the following steps. In step (1) the CO\(_2\) and N\(_2\) adsorption isotherms are predicted using molecular simulations. In step (2), the GCMC adsorption data is used to generate CO\(_2\) purity-recovery Pareto fronts or energy penalty-productivity Pareto fronts using the ANN surrogate model. If the CCS constraints are satisfied for at least one forcefield, then all of the ANN-generated Pareto fronts are refined using the PSA process model. If the CCS constraints are not satisfied for any forcefield combination, then the MOF is discarded from our uncertainty quantification study. In step (3), the hypervolume \( \xi \) of each Pareto front \( i \) and \( j \) is determined by stochastically sampling the Pareto phase space with \( 10^4 \) points, i.e., the blue and red shaded regions in the right-most subplots. The hypervolume error between forcefields \( i \) and \( j \) for a single material at the Pareto level, \( \Delta \xi_{ij} \), is taken as the difference in these two shaded regions, as shown by the grey shaded region in the right-bottom subplot.
At the process level, for each CRAFTED-u MOF, the uncertainty between Pareto fronts that emerges from the use of different molecular FFs to generate the adsorption data as input, is quantified using the hypervolume. The hypervolume, $\xi$, measures the area enclosed by all solutions on the Pareto front and a user-defined reference point $r$.\textsuperscript{77,78} We measure $\xi$ by querying $N = 10^4$ uniformly distributed random points within the Pareto phase space bounded by $r$, where $r$ is determined such that the entire Pareto front for every FF is covered in the sampling. The difference in hypervolumes between FFs $i$ and $j$, which we call the hypervolume error $\Delta \xi^i_j$, is taken as the uncertainty measure for a single material. The process of calculating $\Delta \xi^i_j$ is visually represented in Figure 3.

Now, in order to quantify the global uncertainties between two different FFs $i$ and $j$ across all CRAFTED-u MOFs, we must determine a single numerical value from the full distribution of $\Delta \xi^i_j$ values. We take the mean of the hypervolume error distribution, $\overline{\Delta \xi}^i_j$, as the indicative metric of global uncertainty. Qualitatively we would also like to understand if FF $i$ has a propensity to generate adsorption behaviours that produce better process performances overall compared to FF $j$. Establishing this connection between the adsorption behaviours and process-level responses helps to reveal the biases in performances that may arise in particular FFs. We therefore introduce a normalised metric, $\zeta_i$, to understand the general process-level behaviours of different FFs across many materials which may dominate different areas of KPI phase space. The protocol for calculating $\zeta_i$ for FF $i$ is as follows. For a single MOF, six Pareto fronts are generated for each of the six possible FFs. One of these FFs generates a Pareto front which performs the best, and therefore has the greatest hypervolume. We treat this hypervolume as the upper limit of performance (i.e., $\xi^{max} = 100\%$) and normalise the hypervolume of FF $i$ relative to $\xi^{max}$ such that $\zeta_i = \frac{\xi_i}{\xi^{max}} \times 100$. This operation is applied to every CRAFTED-u MOF, resulting in a distribution of $\zeta_i$ values. Similar to the hypervolume error, we measure the central tendency of the $\zeta_i$ distribution by a single numerical value. In this case, we take the median, $\tilde{\zeta}_i$. If $\tilde{\zeta}_i >$
\( \tilde{r}_j \), then FF \( i \) tends to dominate over FF \( j \) overall. Note that the mean of \( \Delta \tilde{r}_i \) and the median of \( r_i \) were determined to be the most appropriate statistical measures of central tendency given the spread and skewness of the underlying data distributions. They are therefore the best metrics to indicate the overall trends of the CRAFTED-u MOF dataset.

Finally, we measure the correlation between process-level rankings using the Spearman coefficient. Note that, as a matter of convention, we indicate that any of the metrics discussed above are being calculated with respect to a particular FF by the subscript notation and are being compared to another FF using the superscript notation. As an example, \( \tilde{r}_i \) refers to the hypervolume calculated for FF \( i \), and \( \rho_i^j \) refers to the correlation in rankings obtained using FFs \( i \) and \( j \).

3 RESULTS & DISCUSSION

We begin our discussion in section 3.1 by delving into the most crucial and practical consideration of this work, i.e., to what extent the selection of a particular FF impacts the process-level rankings of materials. We consider specific adsorption patterns and the level of agreement between different FFs in predicting the uptake of CO\(_2\) and N\(_2\) in section 3.2. Here, we will reason some of the important differences in molecular modelling approaches, seeking to understand the nature of the (dis)agreement by exploring correlations in simulated CO\(_2\) and N\(_2\) uptakes. In section 3.3 we attempt to establish a connection between the material- and process-levels of description. For this, we will utilise a number of figures and statistical measures to map process-level responses onto the underlying adsorption patterns. Finally, in sections 3.4 and 3.5, we explore potential pathways towards uncertainty mitigation in multiscale simulations of PSA-based CO\(_2\) capture processes and discuss future prospects for HTCS of materials.
3.1 Level of agreement between process-level rankings obtained using different molecular forcefields

In this section, we inquire whether the ranking of porous materials and the identification of top performers through process modelling depends on the use of a specific FF. For this, we must first generate a material ranking list for each molecular FF using some process-level performance metric. The energy-productivity Pareto front, subject to the CCS constraints, is often used to rank materials in multiscale HTCS studies. \(^{63,64,71,75}\) However, the CCS constraints are not satisfied by every CRAFTED-u MOF across all FFs, and so it is not possible to evaluate every material behaviour using energy-productivity KPIs. We therefore ranked materials according to the maximum CO\(_2\) purity they can achieve in the modified Skarstrom cycle (Figure 1), subject to the constraint of CO\(_2\) recovery ≥ 0.9, using the HTCS workflow shown in Figure 3.

Then, we extracted the top 50 MOFs from each ranking, consistent with the philosophy of previous screening studies which typically seek to narrow the candidate list to only the top

![Bar chart showing the number of MOFs in the list of top 50 performing materials shared between the UFF+DDEC forcefield and other molecular forcefields. Materials are ranked by the maximum CO\(_2\) purity value they can achieve in the modified Skarstrom PSA cycle, subject to CO\(_2\) recovery ≥ 0.9. The Spearman coefficient, ρ, measures the strength of the correlation between rankings obtained using the UFF+DDEC forcefield and other molecular forcefields, and is reported above each bar plot.](image)
performers.\textsuperscript{2,16,18,19,71} Upon comparing the materials lists, we find that the different molecular modelling approaches not only do not produce the same rankings, but that they disagree on what materials should be in the top 50. To demonstrate this point, let us make comparisons amongst the different FFs by choosing UFF with DDEC charges as the reference. Note that by doing so, we do not suggest that UFF+DDEC is more accurate in reproducing experimentally observed adsorption behaviours, it is simply a matter of convention. With this in mind, we identify the number of MOFs from each FF-specific ranking which overlap with the top 50 materials from the UFF+DDEC baseline and visualise the results in Figure 4.

Let us first focus on the comparison between UFF and Dreiding in Figure 4, with each of these LJ parameter sets combined with the DDEC charges. We see that only 35 MOFs are shared in the list of top performers between these two variants of the forcefields. Turning now to the combinations of UFF with other charge schemes, the greatest level of agreement is between DDEC and Qeq, with 36 MOFs appearing in both lists of the top 50 materials. It is also quite apparent that the EQeq charges provide a much poorer level of agreement, with only 27 MOFs appearing in the ranking obtained using the DDEC charges. We find that even if we take a different molecular FF as the baseline (Figure S9) or focus on a shorter list of candidate materials, i.e., the top 20 (Figure S10), that generally: (1) there is \textit{consistently} better agreement in the rankings between Qeq and DDEC charges than in EQeq and DDEC charges; (2) depending on one’s choice of the charge scheme, the overlap in the materials identified by any two different schemes ranges between 50\% and 70\% (Figure S9), not to mention the order of materials which changes significantly depending on the set of parameters; (3) the screening results remain sensitive to the choice of LJ parameters and the two lists of top materials produced by two different FFs may share approximately 70\% of the materials. These tendencies are supported by the Spearman ranking correlation coefficients, which are provided above each bar plot in Figure 4.

While Figure 4 is a useful tool to visualise how the identity of the top performers can change with molecular FF, we do not yet understand \textit{how close} the materials are in terms of
their numerical performance. The two possible explanations for moderate-to-poor agreement between material rankings are either: (1) many materials perform similarly, and small differences in the performance estimates give rise to large differences in the relative rankings across different FFs, or (2) the process performances, and therefore the rankings, are highly variable parameters. In Figure 5 we provide a snapshot of how the numerical performance values used to rank the top 10 materials for UFF+DDEC change with molecular FF. From this plot we identify some materials, such as QOVWOO and PEKKAS, which show fairly minor deviations using different FFs, and so are consistently ranked in the top 10 candidates. Other materials, such as VUNCAJ, always remain within the top 50 threshold. Finally, we have some materials with such a strong volatility in their process KPIs that they may fall far below the top 50 threshold, depending on the choice of FF. These constitute the largest proportion of materials, and in many instances the difference in performance is sufficient to classify the same material as suitable (in other words, meeting the CCS constraints) by one FF and not suitable by another FF. Note that these trends persist when the other FFs are taken as the baseline for comparison (Figure S11).
So far it appears that the process performances and material rankings depend on both the definition of the LJ and the electrostatic potentials, however it seems that the choice of charge scheme has a stronger impact (Figure 4). In practice, (E)Qeq may be chosen over DDEC in HTCS\textsuperscript{79–81} due to restricted computational resources, or they may be used to pre-screen large databases and shortlist top performing candidates for further refinement using higher accuracy charge assignment schemes.\textsuperscript{82,83} In an ideal scenario, DDEC charges are available or can be calculated for all candidate materials. In this case, the uncertainty embedded in one’s choice of partial charge assignment would be eliminated as \textit{ab initio} charges are the preferred level of molecular theory for HTCS of materials for carbon capture.\textsuperscript{2} If we confine the scope of our analysis to this ideal scenario, the only additional degree of freedom remaining is how to model the LJ interactions. In Figure 6 (a), a direct comparison of the maximum CO\textsubscript{2} purity predicted by either UFF or Dreiding in combination with DDEC charges reflects this ideal case study. Apart from FUFREE, the MOFs in this top performing subset show similar process-level responses. While it is tempting to attribute this behaviour to minor scattering in the adsorption data, we show later in section 3.4 that this may not strictly be the case as different adsorption patterns can lead to the same CO\textsubscript{2} purity-recovery performances. On the contrary, differences between LJ FFs become much more
pronounced at the energy-productivity Pareto level. For example, in Figure 6 (b) we rank materials according to the maximum productivity they can achieve on their energy-productivity Pareto front (subject to the CCS constraints), and compare how the numerical performance values of the top 10 MOFs identified by UFF+DDEC change when Dreiding+DDEC is used. It is evident from Figure 6 that the predictions between LJ FFs are substantially more uncertain when the energy-productivity process KPIs are used to organise material databases. This observation is notable because the energy-productivity Pareto front more accurately reflects the separation potential of an adsorbent and, unlike the choice of charge scheme, the uncertainty that arises between different LJ parameter sets is unavoidable as the choice of UFF vs Dreiding is still somewhat arbitrary.

What then are the practical implications of these observations for HTCS of materials for CO₂ capture? In light of our results, the preliminary conclusion of this section is that it is very likely that two separate studies utilising different levels of molecular theory will come to different conclusions on what the top performing material candidates are. However, the picture that is emerging cannot be complete without a complementary understanding of the adsorption behaviours which stem from the use of different FFs. Our objective in the next section is to therefore understand the nature of the disagreement between FFs. This will provide us with the molecular insights needed to reason why such differences in process performances arise.

3.2 Analysis of the underlying adsorption patterns and the extent of consistency in CO₂ and N₂ uptakes predicted using different forcefields.

Figure 7 provides a birds-eye-view of the extent of agreement in simulated uptakes of CO₂ and N₂ between different FFs. Figure 7 (a, b) and (e, f) show the Spearman correlations in simulated uptakes for all pressure points. Subplots (a) and (e) show correlations between LJ FFs using fixed charges for CO₂ and N₂, respectively, while subplots (b) and (f) consider fixed LJ FF parameters and different charge schemes. Parity plots of simulated uptakes for
both CO$_2$ and N$_2$ relevant to each of these scenarios are provided in Figure 7 (c, d) and (g, h) at 0.001 bar, i.e., in the Henry’s regime. We begin the discussion by comparing the behaviours of both adsorbates, after which we will explore the differences between molecular FFs. Note that the results presented here are qualitatively similar at higher pressures (see supplementary note 6).

From Figure 7 (a, b) and (e, f), both CO$_2$ and N$_2$ show a greater degree of scattering between different FFs at lower pressures, and therefore poorer correlations. On the other hand, CO$_2$ adsorption shows more scattering compared to N$_2$. We note that CO$_2$ simulated uptakes also seem more sensitive to the electrostatics than to the parameters of the LJ FF, as seen by the stronger correlations in Figure 7 (a) compared to Figure 7 (b); for N$_2$, the opposite trend is observed. The observations above are easy to rationalise. It is known that the most favourable adsorption sites on a MOF will be occupied preferentially at lower pressures.$^{41}$ The strength of the interactions between the adsorbing species and the framework is mediated by the partial charges assigned to framework atoms and the cross-interactions between the adsorbate and adsorbent (and hence the parameters of the LJ FF). Therefore, a higher degree of scattering in the uptakes is expected in the lower pressure regime. CO$_2$ also exhibits more pronounced uncertainties than N$_2$ because: (1) it has more LJ sites available to interact with the MOF, and (2) it has a larger quadrupole moment which strongly interacts with the partial charges assigned to framework atoms.$^{84}$ Finally, as the electrostatic contributions in CO$_2$ adsorption are more significant than in N$_2$ adsorption, the variation in the charge schemes produces a more pronounced scattering in the results compared to variations in the LJ parameters.
Now, we explore in more detail individual adsorption behaviours that contribute to the overall picture in Figure 7. Intuitively, we can expect three primary scenarios for when predictions between FFs deviate significantly from each other: (1) when adsorption occurs in MOFs with small pores; (2) when the LJ parameters for framework atoms are significantly different between UFF and Dreiding; and (3) when charge equilibration methods fail to reproduce DDEC charges. In addition, it is possible that there are combinations of these scenarios for particular MOFs leading to strongly coupled effects. The aim of the following discussion is to explain how these FF deviations arise, and in doing so to provide more insights into the molecular origins of the correlation results in Figure 7 (a, b) and (e, f). To achieve this, we explore each of these three scenarios sequentially, touching on both the general trends and some material specific deviations.

MOFs with smaller pores, and therefore a higher density of framework atoms, are more sensitive to the LJ parameters. These materials are typically associated with lower adsorbate uptakes, and so we generally see higher scattering in the data at lower loadings, as shown in e.g., Figure 7 (c). This is further demonstrated in Figure S12, where the relative
uptake of CO$_2$ at 0.1 bar is plotted against the accessible void fraction of MOFs. In specific cases, confinement effects may also lead to very strong deviations in the adsorptive behaviours depending on the choice of the LJ parameters. For instance, OSOMIT, HELCUY, and GIMVAA – materials which contain Zn and have very narrow pore spaces – represent such exceptional cases. As an example, let us consider the adsorption isotherms of GIMVAA, as shown in the top left panel of Figure 8. The first feature to note about the behaviour of this material is the very strong uptake observed for the UFF+EQeq combination of parameters (green line). This is associated with the EQeq charges assigned to metal atoms in this structure, which are substantially higher than the charges assigned by the DDEC and Qeq schemes. Focussing on the other data in this panel (blue and red lines), we note that only UFF predicts an appreciable uptake of CO$_2$, and much lower uptake for any other combination of parameters. This behaviour stems primarily from the difference in LJ diameters assigned to Zn by UFF and Dreiding ($\sigma_{Zn}^{UFF} = 2.462$ Å and $\sigma_{Zn}^{Dre} = 4.045$ Å). Due to the larger size of the framework atoms, the cross-potential interactions between the adsorbate and adsorbent are unfavourable in the small pores of GIMVAA for Dreiding, and favourable for UFF. This leads to negligible CO$_2$ uptake when using Dreiding-based forcefields.
parameters and non-negligible uptake when the UFF models are employed. Returning to the outlier behaviour of the models based on EQeq charges, GIMVAA is representative of several materials in this class, forming a cluster of green points (cluster I) far above the parity line in Figure 7 (c).

Let us now consider the second scenario, when the LJ parameters for framework atoms are significantly different between UFF and Dreiding. If a MOF contains atom types for which the depth of the well in the interaction potential, $\varepsilon$, between UFF and Dreiding is very different, we can anticipate some sensitivity of the adsorption behaviour to the choice of LJ FF. From Figure 7 (c, g), we see that UFF generally predicts higher adsorbate uptakes. This can be tentatively attributed to the fact that many (approximately 65%) of the CRAFTED-u MOFs contain Zn as the metal which, in the UFF framework, is modelled with $\varepsilon_{\text{Zn}}^{\text{UFF}} = 62.35$ K compared to the Dreiding value of $\varepsilon_{\text{Zn}}^{\text{Dre}} = 27.68$ K. Then, depending on the type of framework atoms and their abundance within the unit cell, it should be possible to determine which MOFs will be sensitive to the choice of LJ FF before performing any molecular simulations. Indeed, Dokur and Keskin developed a simple factor, calculated using the difference in $\varepsilon_i$ between LJ FFs and the number, $n_i$, of atoms $i$, to distinguish MOFs which are relatively insensitive, and MOFs which significantly depend on the choice of the LJ parameters.$^{45}$ For instance, they demonstrated that the deviations between CO$_2$ uptakes predicted by UFF or Dreiding are less than 35% for MOFs with factor values lower than 0.1. Turning our attention now to some specific materials, we find that, similar to the previous scenario, often a particular feature of a MOF tends to magnify the expected deviations between LJ FFs. For example, in addition to OSOMIT, HELCUY, and GIMVAA, cluster I in Figure 7 (c) contains other Zn-metal MOFs such as BEPNIV (Figure 8, top panel on the right) for which confinement effects cannot explain the observed sensitivities as these structures are quite open and feature relatively larger pores. These materials do however contain open metal sites (OMS).$^{65}$ As Zn is defined by a smaller LJ collision diameter in UFF, the closest distance of approach between adsorbing CO$_2$ and the OMS is shorter, which
leads to an exponential increase in the electrostatic contributions to the interactions in these systems. This generally results in more sporadic adsorption behaviours using UFF parameters. From this observation we expect that large uncertainties are likely to be encountered when: (1) MOFs contain OMS that are accessible to the adsorbates; (2) the OMS belongs to a metal in which the difference in atomic diameters between LJ FFs is significant (e.g., Zn, Fe, Ti, Tc, and Ru); and (3) the partial charge assigned to the metal is sufficiently large. Caution is therefore warranted when interpreting the adsorption behaviours in MOFs with OMS using generic LJ FFs, particularly as they often fail to reproduce the adsorption data of their experimental counterparts. Combined, the materials described here and in the previous scenario explain why poorer correlations between LJ FFs are observed using EQeq (Figure 7 (a, e)), and why $\rho_{UUFF+DDEC}^{EQQ} < \rho_{Dre+DDEC}^{DQQ}$ for CO$_2$ in Figure 7 (b).

Finally, the third scenario is concerned with the accuracy of charge assignment using simplified schemes such as Qeq and EQeq. In many instances, charge equilibration methods do not capture the subtle differences in chemical bonding environments that are reflected by DDEC charges. Therefore, only modest agreement between DECC and (E)Qeq charge schemes is observed over the entire CRAFTED-u database. A good example is given in Figure S14, whereby (E)Qeq returns similar charges for all Zn atoms in all CRAFTED-u MOFs, while DDEC charges range between 0 and +1.5$e$ (where $e$ is the elementary charge). For specific materials, charge equilibration methods can fail spectacularly. In particular, we found that the Qeq scheme can return inappropriately high – and in some instances unphysical – charges. Such is the case for the MOFs populating the cluster of blue points (cluster II) below the parity line in Figure 7 (h). The spurious charges assigned in these materials arise for different reasons. For example, in EBOTOF (Figure 8, lower panel on the right), Al metals are coordinated to highly electronegative atoms such as F, which represents a particular bonding environment for which Qeq fails. Similarly, abnormal Qeq charges are assigned in MOFs where Na, Ga, or In metals are present, such
as in AYUWUM (Figure 8, lower left panel). Ongari et al. noticed similar abnormalities in their study.\textsuperscript{44} As to why Qeq fails in this regard, and EQeq does not, stems from the fact that Qeq uses a neutral charge-centre to express the Taylor series expansion in electric energy, while EQeq considers the Taylor series expansion around charge-positive metal cations. The neutral charge-centring is not able to accurately reproduce DDEC charges in alkali metals such as Na, but more generally it appears to fail for MOFs containing metals with a single electron in their outer orbital,\textsuperscript{44} thus leading to similar deviations in In- and Ga-metal MOFs in the CRAFTED-u dataset. These outlier adsorption behaviours, which exist primarily for the Qeq scheme, are what contributes to the difference between Qeq and EQeq correlation lines in Figure 7 (b, f).

At this point, it is clear that the choice of the FF has a very strong impact on the adsorption behaviour, often leading to profound differences in the isotherms for the same material, depending on the parameters used in molecular simulations. Our results further emphasise that using simplified charge equilibration schemes in application to a heterogeneous database of materials is likely to result in unreliable predictions, which can further lead to significant scattering in the results and a lack of confidence in the rankings. From this analysis, we are now better equipped to establish a general connection between the different molecular FFs and the process performance distributions that arise from such modelling choices.

3.3 Quantifying process-level uncertainties emerging from the use of different molecular forcefields

In this section, we quantify the uncertainty between different FFs at the process-level. The objective here is two-fold: (1) we attempt to connect the material uncertainties described in the previous section to the process-level uncertainties in a systematic way; and (2) benchmarking protocols that quantify the impact of different FFs in conditions relevant to the desired application, e.g., carbon capture, will be required as improvements in molecular
simulations continue to develop. Through our analysis, we aim to demonstrate the generality of our framework for evaluating future generations of molecular FFs.

For the above purposes, we make use of the metrics described in section 2.5 to map between distributions in adsorption isotherms and CO$_2$ purity-recovery Pareto fronts. A Pareto-dominance plot, constructed from $\Xi_i$ for each molecular FF $i$, is displayed in Figure 9 (a). This plot demonstrates the tendency of a particular FF to dominate, or not dominate, relative to other FFs. In Figure 9 (b), the process-level uncertainties are visualised using the average hypervolume error, $\overline{\Delta \Xi_i}$. Consistent with the discussion so far, we make comparisons amongst the different FFs by taking UFF+DDEC as the reference. For a more comprehensive comparison of the uncertainties between all pairwise FFs $i$ and $j$, see supplementary note 7.1.

Focussing on Figure 9 (a) first, it is clear that $\Xi_i$ values constructed from the models based on EQeq charges are consistently lower than those using DDEC or Qeq charges. This means that the adsorption behaviours of materials generated using the EQeq charge scheme typically yield poorer process-level performances compared to DDEC and Qeq. Another interesting feature to note is the interaction between LJ FFs and the different charge

![Figure 9. Process-level uncertainty metrics for CO$_2$ purity-recovery optimisations. (a) Pareto-dominance plot constructed from $\Xi_i$ values using the CRAFTED-u dataset. (b) Average hypervolume error, $\overline{\Delta \Xi_i}$, between $i =$ UFF+DDEC and other molecular forcefields ($j \neq$ UFF+DDEC) for the CRAFTED-u dataset. Higher values of $\overline{\Delta \Xi_i}$ indicate a larger average uncertainty with the UFF+DDEC baseline.](image)
schemes. Dreiding yields better average process performances than UFF when coupled with either DDEC or Qeq but poorer average process performances when coupled with EQeq. The molecular origin behind this inverse Pareto-dominance relationship is explained below. Now, turning our attention to Figure 9 (b), we note a number of important trends. Relative to the UFF+DDEC baseline: (1) larger $\Delta \xi_i$ are typically observed when the charge scheme is changed rather than when the LJ FF is changed, i.e., $\Delta \xi^{\text{UFF+DDEC}} > \Delta \xi^{\text{DRE+DDEC}}$; (2) models based on EQeq charges are associated with larger $\Delta \xi_i$ compared to the models based on Qeq charges; and (3) the uncertainties associated with the choice of LJ FF is non-negligible. To expand on this final point, QOVWOO and XOVVIO are the 1st and 47th ranked materials according to the UFF+DDEC forcefield. Their performance curves deviate by a hypervolume of approximately 7%, which is equivalent to the average error between UFF and Dreiding in Figure 9 (b). The results presented here reasonably support the conclusions outlined in section 3.1. In particular, they corroborate that the predictions are overall more sensitive to the choice of the charge scheme than to the choice of the LJ parameters, that the models based on EQeq charges agree less favourably with the DDEC baseline, and that the LJ FF can have a considerable impact on the process-level results.

On the contrary, we find that the connection between the process-level uncertainties and material-level correlations in Figure 7 (a b) and (e, f) is not as straightforward. For example, given that a stronger correlation in the Henry’s regime is observed between DDEC and EQeq in Figure 7 (b, f), one might expect that lower process-level uncertainties in Figure 9 (b) would be observed between DDEC and EQeq rather than DDEC and Qeq. However, different molecular FFs often give rise to variations in several important features of both the CO$_2$ and N$_2$ adsorption isotherms simultaneously, such as the Henry’s coefficient, local slope, nonlinearity, and total saturation capacity. The performance of an adsorbent is determined by an interplay of all of these features and by the competitive adsorption between CO$_2$ and N$_2$. And so it is difficult to capture the process-level effects from a single metric such as the correlation between uptakes at low pressure. This makes
establishing a general connection between molecular FFs and their process-level uncertainties a challenging task. However, it is possible to link individual materials to their process-level responses by a more in-depth analysis. As a brief demonstration, we describe how some of the material behaviours in section 3.2 map to the uncertainty metrics in Figure 9.

Let us revisit the MOFs from cluster II in Figure 7 (h). As a consequence of the high N₂ uptakes predicted in these materials (see for example EBOTOF in Figure S15), the models based on Qeq charges tend to significantly underpredict the process performances relative to those based on DDEC or EEqq charges. This substantial departure from the DDEC baseline leads to very large $\Delta \xi_i^j$ values for Qeq in these materials. However, the existence of these outliers should not suggest that Qeq performs poorer than EEqq overall, as evidenced by the uncertainty metrics in Figure 9 (b) and the material rankings in Figure 4. In fact, our analysis revealed that Qeq – while more prone to sporadic adsorption behaviours, and therefore to outlier process performances – generates $\Delta \xi_i^j$ distributions with lower standard deviations, while EEqq generates $\Delta \xi_i^j$ distributions with greater standard deviations but with fewer statistical outliers (see supplementary note 7.2). So, despite the significant impact of these adsorption behaviours on the correlation metrics, the process-level responses of Qeq are overall more agreeable with DDEC.

Similar to cluster II, the adsorption behaviours of MOFs belonging to cluster I in Figure 7 (c) can be directly linked to the uncertainty results in Figure 9. Taking GIMVAA again as the representative example in this cluster, the UFF+EEqq combination of parameters generates CO₂ isotherms with greater uptakes than the predictions of other FFs (Figure 8, upper left panel). This effect combined with the relatively modest deviations between N₂ isotherms (Figure S15) leads to Pareto fronts which strongly dominate over those generated using other FFs. We therefore see that EEqq in combination with UFF can produce highly favourable process performances such that $\tilde{\gamma}_{UFF+EEqq} > \tilde{\gamma}_{Dreiding+EEqq}$.
while the absence of these behaviours using DDEC or Qeq charges leads to the inverse Pareto-dominance relationship observed in Figure 9 (a). These examples are by no means an extensive account of the individual material-to-process mappings, but they succinctly demonstrate the disparity between material- and process-level metrics.

3.4 Exploring potential pathways towards uncertainty mitigation in multiscale HTCS workflows

In the previous sections we demonstrated that the typical choices and combinations of the FF parameters lead to inconsistent selection of the top performing materials and the relative rankings of the materials. We showed that both selection of the LJ parameters (UFF vs Dreiding) and selection of the charge scheme has an impact on the predictions, however the inconsistency introduced by the charge scheme seems to be more significant compared to the choice of the LJ parameters. Furthermore, our results indicate that charge equilibration schemes may lead to unphysical results (Qeq) and strong deviations in the adsorption behaviours predicted by DDEC charges (both EEq and Qeq).

These findings have significant implications for any further effort to implement fully in silico multiscale screening of porous materials. On the one hand, a strong statement would be that none of the combinations of the parameters are fully validated in experiments and, given the sensitivity of the process predictions to the selection of the FFs, it is unlikely that fully computational workflows will be able to produce reliable results. On the other hand, one could argue that meaningful results can be obtained using this approach provided the parameter inputs and data outputs are handled with care. It is clear that shortcut methods such as (E)Qeq should not be used in HTCS studies, and so future screening studies should adopt ab initio levels of theory for partial charge assignment. The process performance of shortlisted materials identified using ab initio charges should therefore (ideally) be robust against perturbations in the LJ parameters. As a minimum, some estimate of the uncertainty introduced by different LJ parameter sets should be considered to build confidence in the
predictions of these workflows. In either case, an accurate and comprehensive FF is the key bottleneck in computational screening, and we will return to this point in more detail in the next section. For now, we would like to explore two ideas to improve the consistency of predictions within the current state-of-the-art.

The first idea is based on the hypothesis that amongst the CRAFTED MOFs, a group of materials exists whose adsorption behaviours are rather insensitive to the choice of the FF. It is further possible that among these materials, there are those that meet the required purity and recovery constraints and, in general, show good performance at the process level. We speculate that, given the insensitivity of the materials to the FF, it is likely that they will retain their performance in experiments as well. In this case, one can argue that the screening protocols could focus on searching for the materials in this class.

The second idea is as follows. We assert that \textit{ab initio} charges are required for HTCS of MOFs for carbon capture. Unfortunately, it is still computationally challenging to employ these approaches for large databases of materials. Therefore, we would like to explore whether accurate process predictions can be produced using ML models trained to reproduce DDEC-based charges. ML-based charges are an attractive alternative to the DDEC and (E)Qeq schemes as they have the potential to reconcile the benefits of both approaches, i.e., the high accuracy of DDEC and the scalability of (E)Qeq. We anticipate that such models will play a larger role in years to come as the pool of real and hypothetical MOFs continues to expand. We will therefore evaluate the suitability of two state-of-the-art ML-based partial charge assignment models: the \textit{message passing neural network} (MPNN) model\textsuperscript{90} and \textit{partial atomic charges in metal–organic frameworks} (PACMOF) model.\textsuperscript{91}

Let us focus on these ideas in turn. To explore our initial hypothesis that there exists a subset of high performing MOFs that are relatively insensitive to the FF, we first define a taxonomy for describing CRAFTED MOFs by their ability to meet the CCS constraints for different FFs using four distinct consistency classes (Table 2). A small subset of 28 MOFs meets the CO\textsubscript{2} purity and recovery constraints on a consistent basis, regardless of the
particular variation of the FF chosen. We denote these materials as being at the highest level of consistency: Consistency Level 1 (CL1).

<table>
<thead>
<tr>
<th>Class</th>
<th>Label</th>
<th>Number of molecular forcefields (out of 6) for which a material meets the CCS constraints</th>
<th>Number of materials in class</th>
</tr>
</thead>
<tbody>
<tr>
<td>CL1 **</td>
<td>Consistent CCS adsorbents</td>
<td>6</td>
<td>28</td>
</tr>
<tr>
<td>CL2 **</td>
<td>Semi-consistent CCS adsorbents</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>CL3</td>
<td>Inconsistent CCS adsorbents</td>
<td>[1-4]</td>
<td>168</td>
</tr>
<tr>
<td>CL4</td>
<td>Non-CCS adsorbents</td>
<td>0</td>
<td>472</td>
</tr>
</tbody>
</table>

Table 2. Summary of how CRAFTED MOFs are distributed amongst different performance consistency classes.

* The lists of CoRE 2014 DDEC codes for each of these classes are provided in supplementary note 8.
** CL1 and CL2 class materials are differentiated from CL3 and CL4 on the basis of statistical significance. That is, the spread and skewness of the performance distributions for materials in these classes can be visualised using box-and-whisker plots which require a minimum of 5 measurements to construct.

The next question we pose is whether the high level of consistency in the CL1 group is because the adsorption behaviours in these materials are insensitive to the parameters of the FF or due to some other factors. To address this question, in Figure 10 (a) we visualise the relative ranking of CL1 class MOFs, using a box-and-whisker plot to represent the underlying distributions in process performances. It is evident from this subplot that the CL1 class contains materials which are characterised by both narrow distributions in performance, i.e., PEKKAS, and broad distributions in performance, i.e., NOQLOV. This suggests that a search for materials which consistently meet the CCS constraints is not sufficient to guarantee that their performance is insensitive to the choice of FF. Still, within the CL1 subset there are materials such as PEKKAS, QOVWOO, FIRVEH, and NAYZUK, whose narrow boxplots have the potential to confirm our hypothesis. By inspection of the individual adsorption behaviours and performance curves for select materials in Figure 10 (b), we observe a diversity of responses to the choice of FF: the isotherms for some materials depend on both the LJ parameters and charge scheme, like in CUYHIO; in others, it depends mostly on one of these choices, such as in PEKKAS, NOQLOV, and QOVWOO; finally, NAYZUK is the material with the lowest sensitivity to either of these parameters. In either case, some scattering in the adsorption data is observed. This means that the MOFs which are (rather fortuitously) characterised by narrow performance distributions achieve this...
level of consistency through different pathways. That is to say, the CO$_2$ purity-recovery Pareto fronts in these materials converge to similar high performing solutions (despite the scattering seen in the adsorption data) by using very different combinations of the PSA cycle parameters. Therefore, the results in Figure 10 do not provide us with the type of evidence one would like to confirm our initial hypothesis. We arrive then to the conclusion that, in order to achieve consistent process performances in multiscale HTCS workflows for the right reasons, the inconsistencies which propagate from the material level must first be addressed.

![Figure 10](image)

**Figure 10. Relative ranking of CL1 adsorbents under uncertainty.** (a) For each material, a box-and-whisker plot is used to visualise the spread and skewness of the process metrics used to evaluate CL1 class material performances. The process metric is the maximum CO$_2$ purity achievable (subject to the requirement that CO$_2$ recovery $\geq 0.9$). The median process KPI across all forcefield combinations is used to rank materials, as indicated by the solid red line which is provided to guide the eye. Red crosses indicate outlier performance estimates. (b) Distribution of CO$_2$ purity-recovery optimisation results for select materials in class CL1. Top row shows the Pareto fronts for each material, middle row shows the CO$_2$ adsorption isotherms at 298K, and bottom row shows the N$_2$ adsorption isotherms at 298K. Solid lines and dashed lines indicate the UFF and Dreiding forcefields, respectively. Red, green, and blue colours indicate DDEC, EEq, and Eq charges, respectively. Note that the isotherms shown are generated from the dual-site Langmuir model which is used as input for the PSA process simulator.

This brings us onto our second idea. Here, we test whether the uncertainties which are introduced by charge equilibration schemes can be effectively mitigated by using the ML-based MPNN and PACMOF models. As a case study, we continue with the CL1 class of MOFs. These materials satisfy the CCS constraints across all FFs, and so we evaluate the accuracy of MPNN and PACMOF charges by optimising the energy penalty and productivity performances subject to the CCS constraints. This protocol is more robust than an
unconstrained CO₂ purity-recovery optimisation as the energy-productivity process responses are more sensitive to changes in the adsorption behaviours. The optimisations were first conducted using the ANN model and then refined using the PSA process model in order to calculate the hypervolume errors between Pareto fronts generated using different molecular FFs. The results are provided in Figure 11, taking UFF+DDEC as the baseline for comparison.

![Figure 11](image.png)

*Figure 11. Process-level uncertainty metrics for energy penalty-productivity optimisations of CL1 class materials.* Average hypervolume error, $\overline{\Delta \chi_i}$, between $i = UFF+DDEC$ and other molecular forcefields ($j \neq UFF+DDEC$). Higher values of $\overline{\Delta \chi_i}$ indicate a larger average uncertainty with the UFF+DDEC baseline.

Disregarding the MPNN and PACMOF charges in Figure 11 for a moment, we observe a similar response in the energy-productivity results to changes in the FF as the CO₂ purity-recovery results. However, it is clear that scattering in the adsorption data has a stronger influence on these KPIs. For instance, the average hypervolume error of 24% between LJ FFs (combined with DDEC charges) in Figure 11 represents a 3-fold increase in the errors calculated at the CO₂ purity-recovery Pareto level (Figure 9). While these results are determined for a reduced subset of MOFs, we generally expect an uncertainty of approximately 30% between LJ FFs using energy-productivity KPIs (see supplementary note 9). Upon introducing the MPNN and PACMOF charges back into the discussion, the lowest
uncertainties in Figure 11 no longer occur between UFF and Dreiding using fixed DDEC charges. Instead, they occur between DDEC and MPNN / PACMOF charges using fixed UFF parameters, meaning that the prevailing sources of uncertainty identified previously (the choice of charge) has been effectively mitigated by modelling the electrostatics with ML-based charges. Interestingly, the uncertainty that is introduced by the MPNN and PACMOF models is comparable to the scattering seen between surrogate model predictions and PSA process simulations (Figure 2). Efficient pre-screening of MOF databases could therefore be achieved by combining ML models for both the partial charge calculations and the PSA cycle optimisations without a significant loss of accuracy. Overall, our results suggest that MPNN and PACMOF can be used to assign charges in lieu of the DDEC method (see Figures S17, S18, and S20). This is further demonstrated in Figure 12 whereby we see that, of the materials shown here, the adsorption behaviours and process-level predictions obtained using ML-based charges are in closer agreement with the DDEC baseline for all materials apart from NOQLOV.

![Figure 12. Distributions in energy-productivity Pareto fronts for different CL1 class materials. Top row shows the energy-productivity Pareto fronts for each material, middle row shows the CO2 adsorption isotherms at 298K, and bottom row shows the N2 adsorption isotherms at 298K. Note that the isotherms shown are generated from the dual-site Langmuir model which is used as input for the PSA process simulator. From left to right, each subplot shows a material with a lower degree of uncertainty between charge schemes.](image-url)
3.5 Future prospects for HTCS of materials for carbon capture

While we strongly advocate for *ab initio* charges whenever possible, the results presented in Figure 11 and Figure 12 support our argument that ML-based charges are the next best option available to practitioners. Nonetheless, we recognise some of the shortcomings in our analysis (see supplementary note 9) and understand that 28 MOFs cannot represent the diverse chemical landscape of all known MOFs. It is likely that expanding the scope of our study to more materials would expose some regions of poor agreement between ML-based and DDEC charges. Indeed, Liu and Luan\(^{36}\) recently benchmarked several charge assignment methods and determined that charge predictions from PACMOF significantly deteriorated in MOFs which contained transition metals. However, as larger MOF databases with pre-computed *ab initio* charges begin to emerge,\(^ {92}\) we expect that improvements in the current ML models can be achieved by simply extending the training set to cover more materials with greater chemical diversity. Future efforts should therefore be directed towards developing more advanced MOF representations and ML architectures.

A problem still remains when we are posed with the question of which LJ FF to use. Both UFF and Dreiding are designed to be as generic as possible and thus are rather crude approximations of the complex interatomic interactions taking place during adsorption. These interactions are relatively weak compared to the energy variations from chemical bonds, and so the development of more reliable LJ FFs for MOFs remains an open and important challenge. As with most areas of material science, ML has a possible role to play here. The interatomic potentials which describe short range interactions can be constructed using ML methods by approximating the high-dimensional potential energy surface (PES) using training data obtained from *ab initio* calculations.\(^ {93}\) The general and flexible nature of neural network models makes them a viable ML architecture for this purpose. Indeed, the so-called high-dimensional neural network potential (HDNNP) was introduced as early as 2007 by Behler and Parrinello,\(^ {94}\) and in principle considers the PES to be a sum of...
environment-dependent atomic energy contributions. HDNNPs can be computed many orders of magnitude faster than e.g., DFT calculations, they retain the accuracy of reference \textit{ab initio} data, and can be scaled to large systems or time scales.\textsuperscript{93}

These concepts are only beginning to emerge in the field of MOFs. The first application of HDNNPs for MOFs was introduced by Eckhoff and Behler, where they predicted the negative thermal expansion of MOF-5 as well as its phonon density of states using HDNNPs.\textsuperscript{95} Zheng et al. also recently developed a HDNNP to study the chemisorption and diffusion of CO\textsubscript{2} in Mg-MOF-74.\textsuperscript{96} While promising, this approach comes with its own set of open challenges. These include, but are not limited to, the construction of more informative feature vectors which can capture both the local chemical environments and MOF pore attributes, in developing an approach which can model material classes with a large number of chemical elements (the feature vector often scales with the number of chemical elements\textsuperscript{97}), and in generating the reference training data for MOFs with a large number of atoms in their unit cells.\textsuperscript{98} Furthermore, extending this approach to predict adsorption in chemically diverse material databases would require transferable models: the training set must represent the problem at hand. A good example is the general-purpose ANI-2 potential,\textsuperscript{99} which was trained on reference data for a large number of organic molecules and is thus capable of simulating molecules containing (H, C, N, O, F, Cl, and S). One can envision a similar approach being adopted for mixed organic-inorganic complexes. An alternative, less data-intensive avenue would be to use active learning as a means to reliably explore the chemical space of MOFs.\textsuperscript{98,100} In either case, a viable pathway towards developing a general-purpose, highly accurate ML potential for MOF screening applications exists provided that the appropriate training data, ML architecture, and feature vector can be combined. If such a method were to emerge, it has the potential to shift paradigms not only in the screening of MOFs for carbon capture, but in materials modelling in general.
4 CONCLUSIONS

In this multiscale computational screening study, we aimed to address the question: to what extent will the ranking of porous materials, and the selection of top performers identified in process modelling, depend on the choice of molecular forcefield? To do so, we generated distributions of CO₂ and N₂ adsorption isotherms in 690 MOFs using different forcefields (FFs) which represent the modelling choices commonly encountered in high-throughput computational screenings (HTCS) of materials for CO₂ capture. We then conducted a systematic uncertainty quantification study using PSA process-level modelling to determine one’s ability to identify top-performing candidate materials consistently across different FF definitions.

Our results allow us to draw a number of general conclusions:

i) Indeed, the computational ranking of materials appears to depend on the choice of the molecular FF to a significant extent.

ii) From the pool of molecular modelling choices considered in this investigation, the choice of charge assignment scheme represents the largest source of uncertainty at the process-level.

iii) Partial charges assigned by charge equilibration methods such as Qeq and EQeq are unable to reproduce the adsorption behaviours of charges derived from ab initio calculations such as DDEC with sufficient accuracy to guarantee consistent process-level rankings. As such, we recommend avoiding using charge equilibration methods when the electrostatic interactions are important for adsorption, such as CO₂ adsorption in MOFs. When ab initio charges are not available and are computationally unfeasible to obtain, ML-based charges are an attractive alternative that can effectively minimise the uncertainties originating from partial charge assignment. A simple extension in the pool of training materials can provide a quick remedy to the potential pitfalls of existing ML...
models, while future efforts should strive towards advanced MOF representations and machine learning architectures.

iv) We still lack a consistent, experimentally validated set of molecular parameters to describe the LJ interactions in MOFs. Until this issue is addressed, one has to accept the considerable uncertainties embedded within process-level predictions which make use of adsorption data generated using generic LJ FFs.

We believe our work is an important step towards understanding the level of accuracy one can expect from multiscale screenings of materials for PSA-based carbon capture processes. The picture that emerges from our study is that it is not really possible to obtain material rankings with high resolution using the simulation tools currently available. In light of these observations, we see two ways to proceed with HTCS studies.

The first is to strive towards more consistent implementations of these workflows. As we mention above, the most direct route is to only use \textit{ab initio} charges and, failing that, ML-based charges, to model the electrostatic interactions. Moving towards more consistent models of the short-range range dispersion / repulsion interactions would require two stages of implementation, we expect. In the short-term and until a truly universal methodology can be developed, opting for internally consistent LJ parameters within the community is an admirable pursuit. The rational choice would be to use the UFF forcefield, given its ability to simulate MOF databases with greater chemical diversities. However, we still recommend that the performance of shortlisted MOFs should be checked for robustness against perturbations in the LJ parameter sets. In the longer-term, we foresee great possibilities in the use of machine learned interatomic potentials as a means to simulate the short-range interactions in MOFs and, provided such a method can be developed, expect this approach to radicalise the way in which HTCS studies are conducted.

The second option is to alter the perceived utility of HTCS. Many technical barriers exist between identifying MOFs through HTCS and translating them into real, industrially viable separation materials. To name a few, a MOF must demonstrate good mechanical, thermal
and moisture stability, low cost, and scalable synthesisability. A prime example is CALF-20\textsuperscript{101} which, despite its relatively modest process-level performance (see supplementary note 10), is the only known MOF to satisfy enough of these technical requirements to succeed as a viable industrial-scale adsorbent. Therefore, rather than identifying the best performers via material rankings which we understand is not a reliable method to organise databases of MOFs currently, one can instead try to identify good performers using HTCS and then shortlist materials based on all other factors which dictate the feasibility of an adsorbent. Given the importance of these material attributes, this approach could provide more meaning to the results obtained using HTCS and help expedite the transition between identifying promising MOFs and eventually tasking them to bench scale.

Acknowledgements

The authors would like to acknowledge the assistance given by Research IT and the use of the Computational Shared Facility at the University of Manchester.

Code availability

Code will be made available in a GitHub repository to reproduce the results presented in the paper.

Conflicts of interest

There are no conflicts to declare.

5 REFERENCES


59 Non-Linear Least-Squares Minimization and Curve-Fitting for Python — Non-Linear Least-Squares Minimization and Curve-Fitting for Python, https://lmfit.github.io/lmfit-
pyl, (accessed 5 December 2022).


