Electronic Supplementary Material (ESI) for Energy & Environmental Science.

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

(Supporting Information)

Conor Cleeton[1]*, Felipe Lopes Oliveira[2,3], Rodrigo F. Neumann[2], Amir H. Farmahini[1], Binquan Luan[4], Mathias Steiner[2], Lev Sarkisov[1]

[1] The University of Manchester, Manchester, United Kingdom

[2] IBM Research, Av. República do Chile, 330, CEP 20031-170, Rio de Janeiro, RJ, Brazil

[3] Instituto de Qu´ımica, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ, Brazil

[4] 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

Table of Contents

Supplementary Note 1: Additional adsorption simulation details

S1.1 Forcefield parameters used in GCMC simulations of adsorption

The Lennard-Jones (LJ) parameters used to model framework atoms for both the UFF and DREIDING forcefields are provided in Table 1. The Transferable Potentials for Phase Equilibria (TraPPE) forcefield is used to model CO_2 and N_2 gas molecules, and the LJ parameters are provided in Table 3. The critical parameters used to calculate gas phase fugacities with the Peng-Robinson equation of state are provided in Table 2.

Table 1. Lennard-Jones parameters of every framework atom type for UFF and Dreiding forcefields.

Atom type	UFF		Dreiding	
	σ (Å)	ε (K)	σ (Å)	ε (K)
Ca	3.028	119.680	3.093	25.162
Zn	2.462	62.350	4.045	27.678
B	3.638	90.510	3.581	47.808
Al	4.008	253.940	3.911	156.006
Si	3.826	202.150	3.804	156.006
In	3.976	301.210	4.089	276.786
Ga	3.905	208.690	3.911	201.299
Sn	3.913	285.120	3.982	276.786
Ru	2.640	28.160	4.044	27.6786
Ge	3.813	190.580	3.804	201.299
Tc	2.671	24.140	4.044	27.678
Sb	3.938	225.780	3.875	276.786
As	3.769	155.380	3.697	206.331
Te	3.982	200.140	3.768	286.851
Br	3.732	126.220	3.519	186.201
Ti	2.829	8.550	4.045	27.678
Fe	2.594	6.540	4.045	27.678
Na	2.658	15.090	2.801	251.623
H	2.571	22.140	2.846	7.649
C	3.431	52.800	3.472	47.858
P	3.695	153.370	3.697	161.039
N	3.261	34.700	3.262	38.951
Cl	3.516	114.150	3.519	142.570
S	3.595	137.780	3.590	173.117
F	2.997	25.140	3.093	36.485
Se	3.746	146.330	3.590	216.396
O	3.118	30.200	3.033	48.160

Table 2. Critical parameters for CO_2 and N_2 used in the calculation of gas-phase fugacities

Gas	Critical Temperature (K)	Critical Pressure (Pa)	Acentric Factor
CO_2	304.1282	7377300.0	0.22394
N_2	126.192	3395800.0	0.0372

Table 3. Lennard-Jones Parameters for TraPPE forcefield

Atom type	σ (Å)	ε (K)	Charge
C_co2	2.80	27.00	0.700
O_co2	3.05	79.00	-0.350
N_n2	3.31	36.00	-0.482
N_com	-	-	0.964

S1.2 Additional notes on GCMC simulation protocol

Simulations of CO_2 and N_2 adsorption are performed using the atomistic GCMC method implemented in RASPA [1]. Interaction energies between non-bonded atoms are modelled using the LJ plus Coulomb potential (Eq. (7)):

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

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \qquad \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \quad (2)$$

Where i and j are indices of the interacting atoms, r_{ij} is the distance between atoms i and j, ε and σ are the LJ parameters, ε_0 is the vacuum permittivity, and q is the partial charge. 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. The Ewald summation technique was used to calculate the long-range electrostatic interactions (for Qeq, EQeq, and DDEC partial charges) with a relative precision of 10^{-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.

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) [2] 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. 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.

S1.3 Validation of competitive dual-site Langmuir model

The process model described in the main article, section 2.3, takes as input dual-site Langmuir (DSL) model parameters (regressed against single component GCMC simulated isotherms) to make predictions on the competitive adsorption of CO_2 / N_2 mixtures. We impose constraints during the DSL model fitting process (described in section 2.2 of the main article) to obtain physically meaningful parameters. As validation, we compared the competitive DSL model predictions to the simulated uptakes of binary CO_2 / N_2 gas mixtures using GCMC (Figure S1) for four MOFs randomly chosen from the CRAFTED (v1.1.1) database. The competitive DSL model predictions are in good agreement with the GCMC simulation results up to 5 bar, which is the maximum pressure allowed in the adsorption step of the pressure swing adsorption cycle.

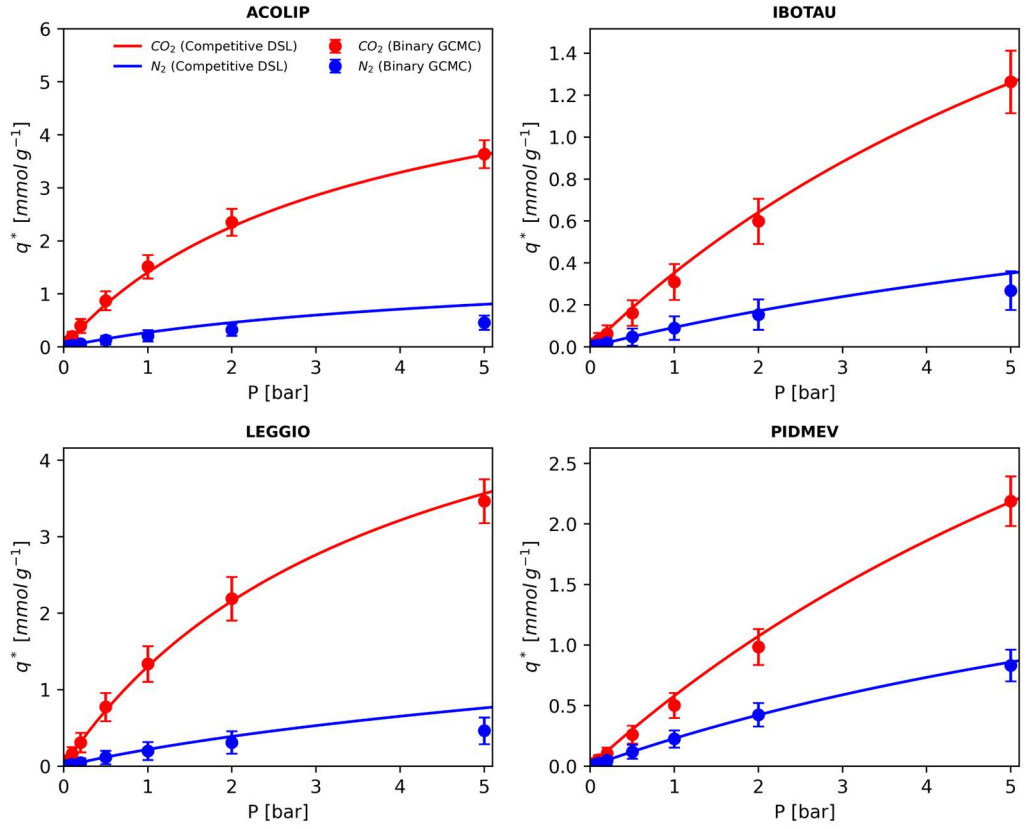

Figure S1. **Loading as a function of the total pressure for a 15% CO_2 / 85% N_2 mixture at 315.15K (i.e., at conditions representative of the adsorption step in the modified Skarstrom PSA cycle).** *Binary GCMC simulations results given by the circles with error bars, and the predictions by competitive dual-site Langmuir models fit using single component GCMC data are shown by the solid lines. Four MOFs, namely ACOLIP, IBOTAU, LEGGIO, and PIDMEV, are chosen at random from the CRAFTED database for validation.*

Supplementary Note 2: Additional PSA simulation details

This supplementary note provides additional details on the simulation protocol used to model the modified Skarstrom cycle detailed in section 2.3 of the main article. The non-dimensionalised conservation equations and associated boundary conditions are given in Table 4. Mathematical definitions of the 4 key performance indicators (KPIs) are given by Eqs. (3) - (6). In Table 5, the complete list of inputs needed to solve the conservation equations detailed in Table 4 is provided.

Table 4. Mathematical model and boundary conditions for different steps in the modified Skarstrom cycle.

Nondimensional Variables:

$$\bar{p} = \frac{p}{p_0};\ \bar{T} = \frac{T}{T_0};\ x_i = \frac{q_i}{q_{s0}};\ x_i^* = \frac{q_i^*}{q_{s0}};\ \bar{v} = \frac{v}{v_0};\ \bar{z} = \frac{z}{L};\ \bar{t} = \frac{t \cdot v}{L}$$

Pellet (adsorbed phase) Mass Balance:

$$\frac{\partial x_i}{\partial \bar{t}} = \frac{k_i^{LDF} L}{v_0}(x_i^* - x_i); \qquad k_i^{LDF} = \frac{c_i}{q_i^*}\frac{15}{r_p^2}\frac{\varepsilon_p}{\tau} D_{macro,i}$$

Component Mass Balance:

$$\frac{\partial y_i}{\partial \bar{t}} = \frac{1}{Pe}\left(\frac{\partial^2 y_i}{\partial \bar{z}^2} + \frac{1}{\bar{p}}\frac{\partial \bar{p}}{\partial \bar{z}}\frac{\partial y_i}{\partial \bar{z}} - \frac{1}{\bar{T}}\frac{\partial y_i}{\partial \bar{z}}\frac{\partial \bar{T}}{\partial \bar{z}}\right) - \bar{v}\frac{\partial y_i}{\partial \bar{z}} + \frac{\Psi \bar{T}}{\bar{p}}\sum_{i=1} y_i \frac{\partial x_i}{\partial \bar{t}}$$

Total Mass Balance:

$$\frac{\partial \bar{p}}{\partial \bar{t}} = -\frac{\partial}{\partial Z}\left(\frac{\bar{v}\bar{p}}{\bar{T}}\right) - \Psi \bar{T}\sum_{i=1}\frac{\partial x_i}{\partial \bar{t}} + \frac{\bar{p}}{\bar{T}}\frac{\partial \bar{T}}{\partial \bar{t}}$$

Energy Balance:

$$\left(\varepsilon_B c_{p,g}\rho_g + (1-\varepsilon_B)(\rho_p c_{p,s} + c_{p,a} q_{s0})\right)\frac{\partial \bar{T}}{\partial \bar{t}} = \frac{K_z}{v_0 L}\frac{\partial^2 \bar{T}}{\partial \bar{z}^2} - \varepsilon_B \rho_g c_{p,g}\bar{v}\frac{\partial \bar{T}}{\partial \bar{z}} + \sum_{i=1}(1-\varepsilon_B)(-\Delta H_i^{st})\frac{q_{s0}}{T_0}\frac{\partial x_i}{\partial \bar{t}}$$

Initial Conditions during Start-up:

$$\bar{p} = \overline{p_L};\quad y_i = y_{i,feed};\quad \bar{T} = \bar{T}_a;\quad x_i = x_i^*|_{y_{feed}}$$

Cyclic steady state condition:

$$||\mathrm{X}|| = \sqrt{\sum_{k=1}^{N}|\mathrm{X_k}|^2} < 10^{-4} \qquad \underline{\text{OR}} \qquad no.\,cycle\ iterations = 700^*$$

Boundary Conditions	Pressure (p)	Mole frac. (y_i)	Temp. (T)	
(1) *Pressurisation*	$\bar{p} = \overline{p_L} \rightarrow 1^{**}$	$y_i = y_{i,feed}$	$\bar{T} = 1$	Bottom
	$\frac{\partial \bar{p}}{\partial \bar{z}} = 0$	$\frac{\partial y_i}{\partial \bar{z}} = 0$	$\frac{\partial \bar{T}}{\partial \bar{z}} = 0$	Top
(2) *Adsorption*	$\bar{p} = 1.02$	$y_i = y_{i,flue}$	$\bar{T} = \bar{T}_{flue}$	Bottom
	$\bar{p} = 1$	$\frac{\partial y_i}{\partial \bar{z}} = 0$	$\frac{\partial \bar{T}}{\partial \bar{z}} = 0$	Top
(3) *Heavy Reflux*	$\bar{p} = 1.02$	$y_i = y_{i,LR}\|_{\bar{z}=0}$	$\bar{T} = \bar{T}_{LR}\|_{\bar{z}=0}$	Bottom
	$\bar{p} = 1$	$\frac{\partial y_i}{\partial \bar{z}} = 0$	$\frac{\partial \bar{T}}{\partial \bar{z}} = 0$	Top
(4) *CnC-Depressurisation*	$\bar{p} = 1 \rightarrow \overline{p_L}$	$\frac{\partial y_i}{\partial \bar{z}} = 0$	$\frac{\partial \bar{T}}{\partial \bar{z}} = 0$	Bottom
	$\frac{\partial \bar{p}}{\partial \bar{z}} = 0$	$\frac{\partial y_i}{\partial \bar{z}} = 0$	$\frac{\partial \bar{T}}{\partial \bar{z}} = 0$	Top
(5) *Light Reflux*	$\bar{p} = \overline{p_L}$	$\frac{\partial y_i}{\partial \bar{z}} = 0$	$\frac{\partial \bar{T}}{\partial \bar{z}} = 0$	Bottom
	$\bar{p} > \overline{p_L}$	$y_i = y_{i,ads}\|_{\bar{z}=1}$	$\bar{T} = \bar{T}_{ads}\|_{\bar{z}=1}$	Top

Additional Info.

$p_0 = p_H;\ T_0 = 313.15K$

$q_{s0} = 5.84\frac{mol}{kg};\ L = 1$

$v_0 = f(p_0, T_0, v_{feed})$

$D^L = 0.7D_m + v_0 r_p$

$Pe = \frac{v_0 L}{D^L}$

$\Psi = \frac{RT_0 q_{s0}}{p_0}\frac{(1-\varepsilon_B)}{\varepsilon_B}$

$\rho_p = (1-\varepsilon_p)\rho_s$

X_k is the difference in state variable k between the final conditions of the light reflux step and the initial conditions of the pressurisation step.

* From our preliminary tests of the simulation code we confirm that the number of cycle iterations generally do not exceed 700.

** Pressure at the inlet or outlet of the column is calculated using an exponential function of the form: $P = P_{set} + (P_{set} - P_{init})e^{-\lambda t}$ where λ = 0.5 for PSA cycle steps where the pressure in the column is changing.

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 [3] which relies on the simulation code of Yancy-Caballero et al. [4] 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 [5]. A "unibed" simulation strategy is adopted [6], 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.

At CSS the performance of the cycle configuration is evaluated using four descriptors, namely the purity of CO_2 [–], recovery of CO_2 [–], energy penalty [$kWh\ ton_{CO_2}^{-1}$], and productivity [$mol_{CO_2}\ kg^{-1}s^{-1}$] of the cycle, defined in Eqs. (3) - (6).

$$Purity_{CO_2} = \left(n_{CO_2}^{CnCDepres} + (1-\alpha_{HR})n_{CO_2}^{LR}\right)/\left(n_{total}^{CnCDepres} + (1-\alpha_{HR})n_{total}^{LR}\right) \quad (3)$$

$$Recovery_{CO_2} = \left(n_{CO_2}^{CnCDepres} + (1-\alpha_{HR})n_{CO_2}^{LR}\right)/\left(n_{CO_2}^{Pres} + n_{CO_2}^{Ads}\right) \quad (4)$$

$$Energy\ Penalty = \sum_{step}^{N_{steps}} E_{step}/\left(n_{CO_2}^{CnCDepres} + (1-\alpha_{HR})n_{CO_2}^{LR}\right) \cdot MW_{CO_2} \cdot 10^{-6} \quad (5)$$

$$Productivity\ = \left(n_{CO_2}^{CnCDepres} + (1-\alpha_{HR})n_{CO_2}^{LR}\right)/m_s \cdot \sum_{step}^{N_{steps}} t_{step} \quad (6)$$

Where $n_{CO_2}^{step}$ and n_{total}^{step} $[mol]$ are the CO_2 and total molar quantities in a given step, m_s $[kg]$ is the mass of the solid adsorbent, MW_{CO_2} $[g\ mol^{-1}]$ is the molecular weight of CO_2, t_{step} $[s]$ is the time of a given step, and E_{step} $[kWh]$ is the mechanical energy associated with pressurisation or pulling vacuum for a given step. The power for the pumps / vacuum pumps is given by Eq. (7):

$$E_{step} = \begin{cases} \frac{1}{\eta}\varepsilon_B \cdot \pi \cdot r_{in}^2 \cdot \frac{\gamma}{\gamma-1} \int_0^{t_{step}} (v \cdot p|_{z=0}) \left[\left(\frac{p|_{z=0}}{p_{atm}}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] dt\,, & if\ p|_{z=0} \geq p_{atm} \\ \frac{1}{\eta}\varepsilon_B \cdot \pi \cdot r_{in}^2 \cdot \frac{\gamma}{\gamma-1} \int_0^{t_{step}} (v \cdot p|_{z=0}) \left[\left(\frac{p_{atm}}{p|_{z=0}}\right)^{\frac{\gamma-1}{\gamma}} - 1\right] dt, & if\ p|_{z=0} \leq p_{atm} \\ 0 \quad , & otherwise \end{cases} \quad (7)$$

Here, ε [–] is the void fraction of the packed bed, r_{in} $[m]$ is the inner radius of the adsorption column, v is the interstitial velocity $[m\ s^{-1}]$, p_{atm} and $p|_{z=0}$ $[Pa]$ are the atmospheric pressure and pressure evaluated at the feed end of the column, η [–] is the isentropic efficiency of the pumps / vacuum pumps, which is assumed to be 72% [7], and γ [–] is the adiabatic constant.

Table 5. Complete list of input parameres used to model the modified Skarstrom cycle.

Column Properties	**Value**
Length of column, $L\ [m]$	1
Radius of column, $r_B\ [m]$	0.1445
Flue Gas Parameters and Constants	
Universal gas constant, $R\ [J\ mol^{-1}K^{-1}]$	8.314
Feed Temperature, $T_{feed}\ [K]$	313.15
Feed CO_2 mole fraction, $y_{CO_2}\ [-]$	0.15
Fluid viscosity, $\mu\ [kg\ m^{-1}s^{-1}]$	1.72×10^{-5}
Effective gas thermal conductivity, $K_z\ [W\ m^{-1}K^{-1}]$	0.09
Specific heat capacity of adsorbed phase, $c_{p,a}\ [J\ kg^{-1}K^{-1}]$	30.7
Specific heat capacity of gas phase, $c_{p,g}\ [J\ kg^{-1}K^{-1}]$	30.7
Transport Properties	
Mass transfer coefficient of CO_2, $k_{CO_2}\ [s^{-1}]$	Variable*
Mass transfer coefficient of N_2, $k_{N_2}\ [s^{-1}]$	Variable*
Molecular diffusivity, $D_m\ [m^2s^{-1}]$	1.2995×10^{-5}
Operating Conditions	
Low pressure, $p_L\ [bar]$	0.1
Time of pressurisation step, $t_{pres}\ [s]$	20
Time of heavy reflux step, $t_{HR}\ [s]$	t_{ads}
Time of evacuation step, $t_{evac}\ [s]$	15
Time of light reflux step, $t_{LR}\ [s]$	t_{ads}
Pressure correction parameter, $\tau\ [-]$	0.5
Constants for dimensionless parameters	
$p_0\ [bar]$	P_{feed}
$T_0\ [K]$	T_{feed}
$v_0\ [m\ s^{-1}]$	v_{feed}
$q_{s0}\ [mol\ kg^{-1}]$	5.84
Adsorbent Parameters	
Crystal density, $\rho_s\ [kg\ m^{-3}]$	600 – 2600
Specific heat capacity of adsorbent, $c_{p,s}\ [J\ kg^{-1}K^{-1}]$	500 – 1500
Pellet tortuosity factor, $\tau'\ [-]$	3
Adsorption saturation capacities, $q_{s,j}^{i}\ [mmol\ g^{-1}]$	0 – 12.5 **
Pre-exponential factors, $b_{0,j}^{i}\ [bar^{-1}]$	10^{-9} – 10^{-2} **
Heat of adsorption for CO_2, $\Delta H_{CO_2,j}^{ads}\ [J\ mol^{-1}]$	(-15000) – (-45000) **
Heat of adsorption for N_2, $\Delta H_{N_2,j}^{ads}\ [J\ mol^{-1}]$	(-15000) – (-45000) **
Optimisable Parameters	
Flue gas adsorption pressure, $P_{feed}\ [bar]$	1 – 5
Time of adsorption step, $t_{ads}\ [s]$	5 – 500
Interstitial feed velocity, $v_{feed}\ [m\ s^{-1}]$	0.1 – 2
Light reflux ratio, $\alpha_{LR}\ [-]$	0.01 – 0.2
Bed void fraction, $\varepsilon_B\ [-]$	0.35 – 0.45
Radius of pellet, $r_p\ [m]$	0.5×10^{-3} – 2.5×10^{-3}
Pellet void fraction, $\varepsilon_p\ [-]$	0.3 – 0.7

* Mass transfer coefficient is dependent on the isotherm of a given adsorbent and is determined using the k_i^{LDF} equation in Table 4.

** i = component CO_2 or N_2; j = site 1 or site 2 of the dual-site Langmuir model.

Supplementary Note 3: Sensitivity analysis and dimensionality reduction

A preliminary sensitivity analysis of the energy penalty-productivity Pareto front was conducted to determine the importance of PSA cycle variables. The performance of 15 hypothetical adsorbents, characterised by random combinations of material properties (heat capacities, crystal densities, and DSL model parameters), were optimised using the high-fidelity PSA process model and NSGA-II algorithm. This problem can be formulated as:

$$\Phi_{min} = \min_{\vartheta} \Phi_i \quad i = \{1,2\} \tag{8}$$
$$\Phi_1 = Energy\ Penalty$$
$$\Phi_2 = -Productivity$$
$$s.t.: \quad lb < \vartheta = [p_H, p_L, t_{pres}, t_{ads}, t_{CnCDepres}, \alpha_{LR}, \alpha_{HR}, v_{feed}] < ub$$
$$Purity_{CO_2} > 0.9$$
$$Recovery_{CO_2} > 0.9$$

Here, ϑ is the vector of decision variables which we are seeking to optimise. After confirming that each Pareto front converged, the optimal distribution of each PSA cycle variable was visualised to determine which are important and which are not. It must be highlighted that a 'non-important' or 'non-sensitive' variable in this context does not suggest that perturbations in this parameter leads to a negligible response in the predicted KPIs. Rather, it alludes to the fact that at the Pareto level this parameter converges to similar values, and thus does not represent a truly optimisable parameter. To illustrate this, consider Figure S2 where the variable distributions of 'non-important' parameters - p_L, α_{HR}, t_{pres}, and $t_{CncDepres}$ (t_{evac}) – are plotted for the 15 hypothetical energy-productivity Pareto fronts. The evacuation

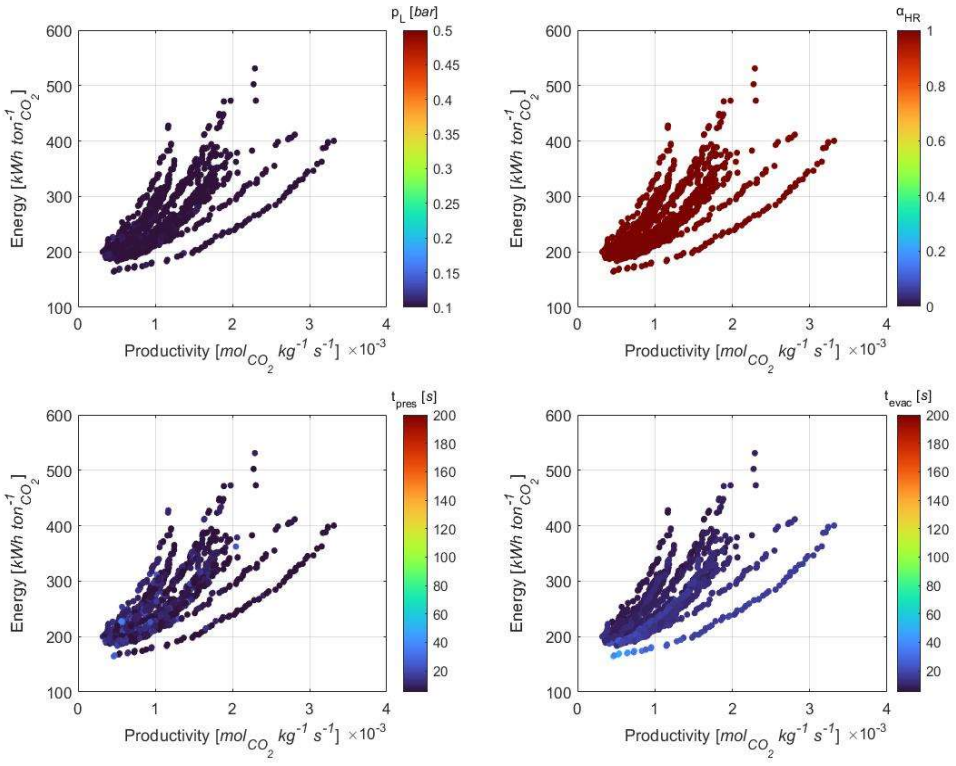

Figure S2. **Energy-productivity Pareto fronts for 15 hypothetical adsorbent materials coloured according to their 'non-important' variable distributions.**

pressure, p_L, is known to strongly influence the performance of individual cycle configurations [8,9], however at the Pareto-level the optimal value is consistently determined to be at the lower limit of 0.1 bar. As such, the search space can be reduced by simply fixing the evacuation pressure to 0.1 bar. Similar trends can be observed with the rest of the 'non-important' variables in Figure S2. In contrast, the variable distributions for p_H, t_{ads}, v_{feed}, and α_{LR} in Figure S3. are deemed to be 'important', and thus are retained in subsequent process-level optimisations.

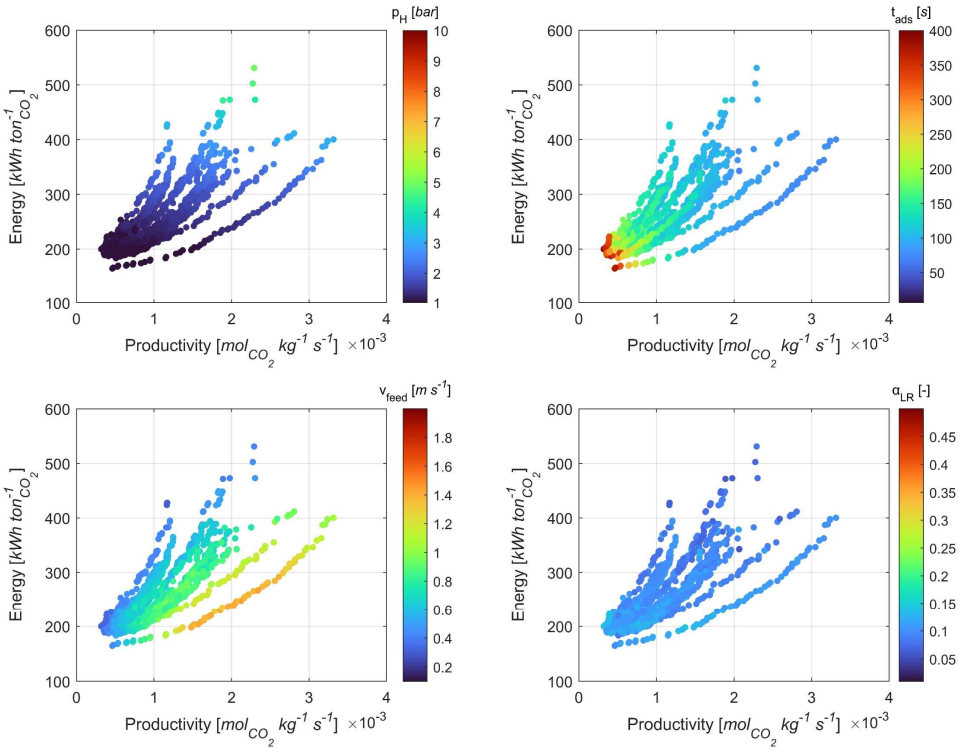

Figure S3. **Energy-productivity Pareto fronts for 15 hypothetical adsorbent materials coloured according to their 'important' variable distributions.**

Supplementary Note 4: Additional notes on surrogate modelling protocol

S4.1 Material property ranges in the CRAFTED MOF dataset

Appropriate training limits must be defined for the surrogate model such that the performance of all CRAFTED MOFs can be predicted. The material-specific parameters are the crystal specific heat capacity $c_{p,s}$ $[J\ kg^{-1}K^{-1}]$, crystal density ρ_s $[kg\ m^{-3}]$, and the CO_2 and N_2 adsorption isotherms. In Figure S4, we show the distributions in material properties for the full CRAFTED MOF dataset. The specific heat capacities were calculated using the machine learning model of Moosavi et al. [10], while crystal densities (and other structural properties) were calculated using Zeo++ [11] with a probe radius of 1.655 Å and high accuracy settings. The adsorption isotherm calculations are described in section 2.1 of the main article.

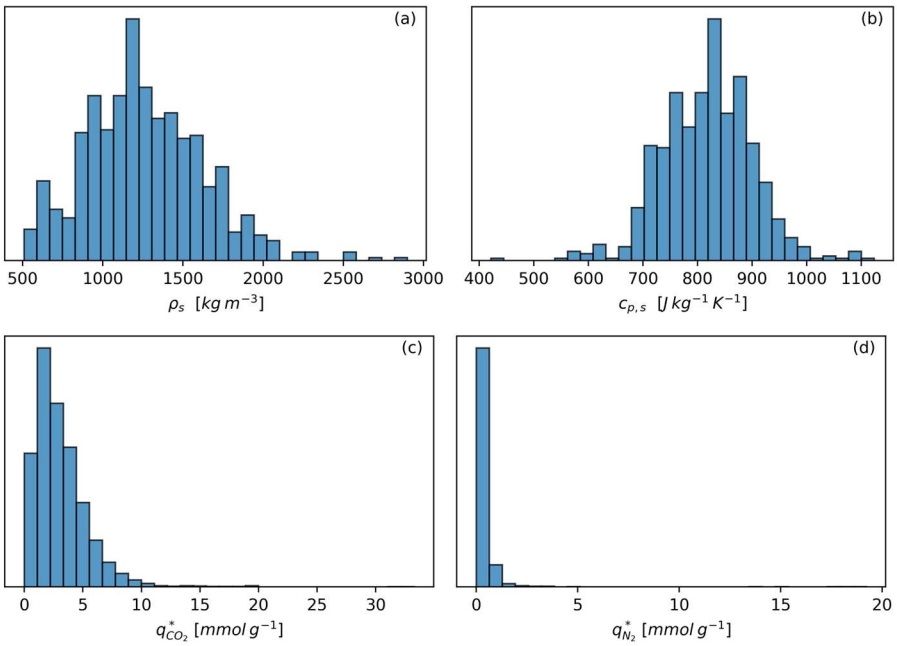

Figure S4. **Distribution of CRAFTED MOF material properties.** (a) Distribution of crystal densities. (b) Distribution of crystal specific heat capacities. (c) Distribution of CO_2 uptakes for all CRAFTED MOFs, using six different molecular forcefields (described in the main article, section 2.1), at 298K and 1 bar. (d) Distribution of N_2 uptakes for all CRAFTED MOFs, using six different molecular forcefields (described in the main article, section 2.1), at 298K and 1 bar.

S4.2 ANN model training and refinement

ANN Model Training

There are overall 7 optimisable parameters (described the main article, section 2.3 and 2.4) and 14 material-specific parameters ($c_{p,s}$, ρ_s, and the 12 DSL parameters for 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). With the material property ranges defined, we first generated 100,000 unique combinations of the 21 inputs using the Latin hypercube sampling (LHS) technique and generated their associated process KPIs using the high-fidelity PSA process model. We then cleaned and pre-processed the training data using the following criteria:

- We removed any samples which generated a CO_2 purity lower than 0.15 or produced a CO_2 purity / recovery > 1.005 (i.e., the max CO_2 purity / recovery possible plus the mass balance error tolerance). These cycle configurations did not satisfy the mass balance requirements and are thus discarded.
- We removed samples which exceeded 700 cycle iterations when determining the cyclic steady state (CSS) solutions. This number of cycle iterations was an additional exit condition encoded into the PSA simulations to ensure that any individual cycle configuration did not take too long to compute. Generally, this exit condition was rarely invoked. However, we cannot guarantee that CSS was achieved in these configurations and so they were discarded from the training set.
- We removed outlier datapoints that produced energy penalties greater than 10^4 $kWh\ ton_{CO_2}^{-1}$ and productivities greater than 0.01 $mol_{CO_2}\ kg^{-1}s^{-1}$.
- We were interested in maximising the CO_2 purity and recovery for a given material and forcefield, as described in the main article. These properties are physically limited between 0 and 1 ± 0.005. ANN model predictions can violate these physical constraints as the underlying physics of the process is not explicitly encoded into the model architecture. Therefore, to address this issue we instead train the ANN on transformed purity and recovery responses of the form given in Eq. (9), while real-line predictions are made by transforming back using Eq. (10) [12]:

$$\hat{y} = \ln\left(\frac{y}{(1+\epsilon) - y}\right) \tag{9}$$

$$y = \frac{(1+\epsilon)\exp\hat{y}}{1+\exp\hat{y}} \tag{10}$$

  Here, $\hat{y}$ is the transformed response, y is the original output, and ϵ is a small positive perturbation to allow y to be equal to 1 ± 0.005. This transformation preserves the Pareto-ordering of the objective functions and confines the ANN predictions to physically meaningful values when coupled with the NSGA-II optimiser.
- The energy penalty and productivity distributions were highly skewed, and so to improve the ANN model predictions we trained on a log-like transformation of the form given in in Eq. (11), while real-line predictions are made by transforming back using Eq. (12):

$$\hat{y} = \frac{\ln(y)}{\ln(10)} \tag{11}$$

$$y = 10^{\hat{y}} \tag{12}$$

After applying these pre-processing operations, we trained an individual ANN model for each of the 4 KPIs. In our previous work [3] we found that, after an extensive hyperparameter optimisation, that ANN models with 3 hidden layers and 20 neurons per hidden layer were sufficient to achieve good predictions on samples unseen during the model training. Upon extending the material-specific parameter ranges in the current work – which was necessary to include the full distribution of CRAFTED MOF material properties – we found that more neurons were required in each hidden layer to achieve the same level of accuracy. We opted for 45 neurons per hidden layer. Each ANN model was trained using the Levenberg-Marquardt back-propagation technique for 2500 epochs using 90% of the training data, or until the validation mean squared error loss indicated overfitting. This process was repeated 5 times using different test sets in each iteration to obtain the lowest root mean squared prediction error (RMSE) for every ANN model.

ANN model refinement

100,000 samples were indeed sufficient to train a highly accurate surrogate model. However, using only the LHS scheme, it was very difficult to obtain high productivity values, meaning that an ANN conditioned on this dataset would perform poorly in the Pareto-optimal energy-productivity phase space of high performing materials. We therefore supplemented the original 100,000 data points with high quality, Pareto-optimal points that were obtained using a bootstrap optimisation technique. The details of this approach are as follows. The high-fidelity PSA process model was coupled with the NSGA-II optimisation algorithm, after which we perform multiple optimisations of the energy penalty and productivity using initial populations that were generated using different random seeds. In these optimisations, the PSA cycle variables and pellet properties were optimised, while the material properties

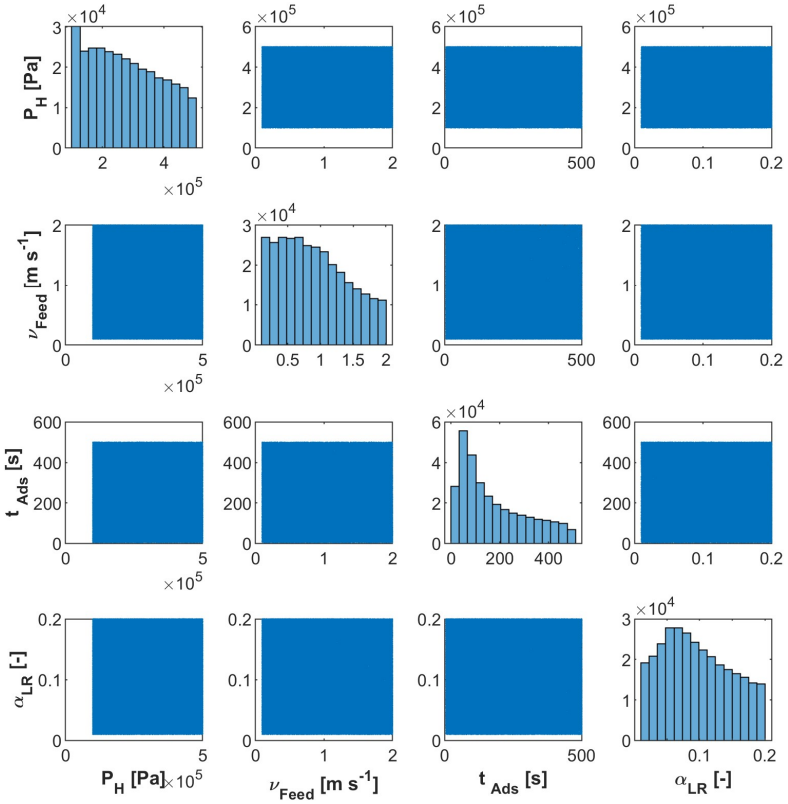

Figure S5. **Distribution of 300,000+ samples of the PSA cycle variables used to train the artificial neural network model.** The visualisation shows that the Latin hypercube sampling and bootstrap optimisation sampling sufficiently covers the entire input phase space.

($\rho_s, c_{p,s}$ DSL model parameters) were mutated with each successive generation to ensure a diverse set of both the design variables and material properties were sampled. We extracted the unique combinations of inputs that produce favourable energy penalty-productivity performances and added them to the pool of training data. Overall, this combination of LHS and guided-search sampling generated over 300,000 unique training points, which we used to train the final ANN model. The ability of the ANN model to predict both individual cycle configurations and fully resolved Pareto fronts is demonstrated in the main article, section 2.4. In Figure S5 – Figure S7, the distribution of inputs used in the training are shown. In Figure S8, the distribution of outputs used in the training are shown.

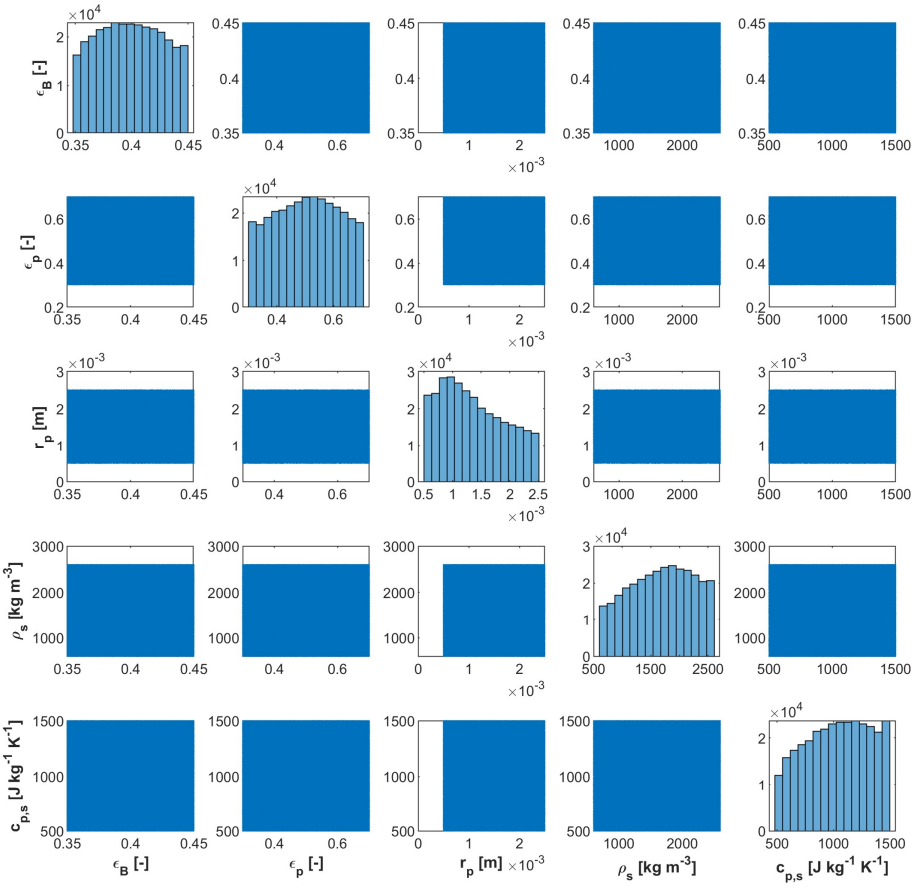

Figure S6. **Distribution of 300,000+ samples of the material-specific and pellet-specific variables used to train the artificial neural network model.** The visualisation shows that the Latin hypercube sampling and bootstrap optimisation sampling sufficiently covers the entire input phase space.

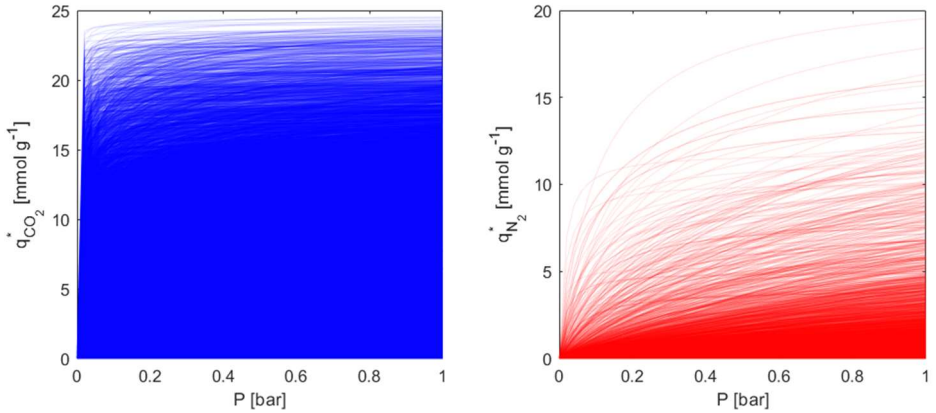

Figure S7. **Distribution of isotherms generated from hypothetical samples of the DSL model parameters used to train the artificial neural network model.** Note, only 100,000 samples are shown here as demonstration, however over 300,000 samples are used in the training.

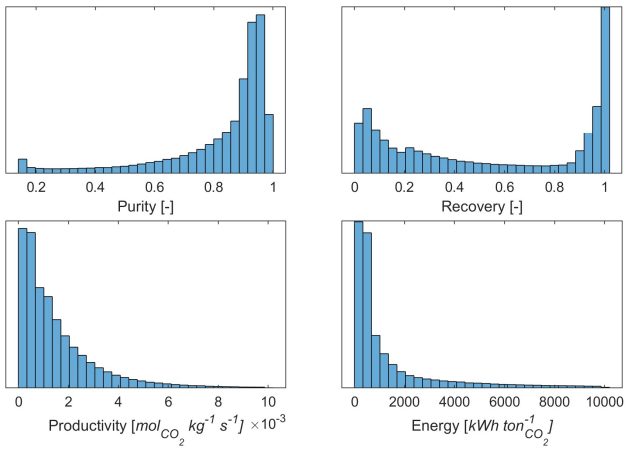

Figure S8. **Distributions of target KPIs used in the training of the artificial neural network model.**

Supplementary Note 5: Additional figures for section 3.1 of the main article

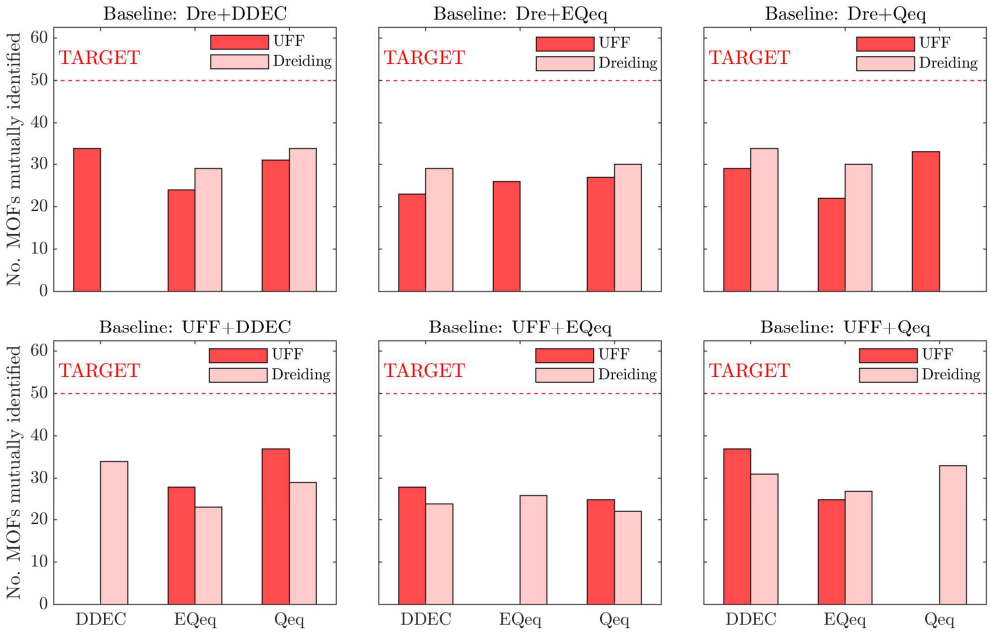

Figure S9. **Identifying the number of top 50 MOFs ranked by each molecular forcefield (title of each subplot) which overlap with the rest of the molecular forcefields**. Rankings of materials are organised by the maximum CO_2 purity value they can achieve in the modified Skarstrom PSA cycle, subject to CO_2 recovery ≥ 0.9, using different molecular forcefields.

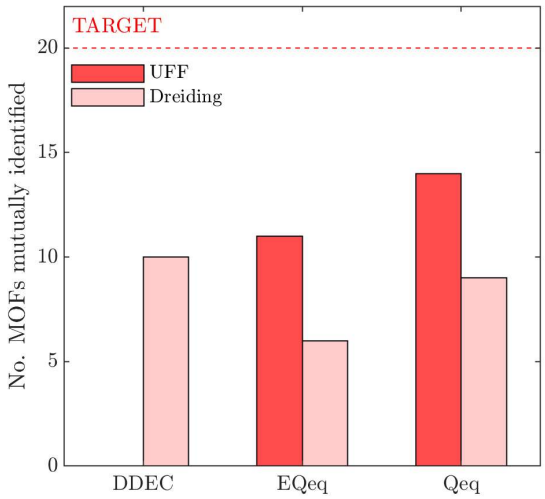

Figure S10. **Number of the top 20 MOFs ranked by UFF+DDEC which are mutually identified by other molecular forcefields.** Rankings of materials are organised by the maximum CO_2 purity value they can achieve in the modified Skarstrom PSA cycle, subject to CO_2 recovery ≥ 0.9, using different molecular forcefields.

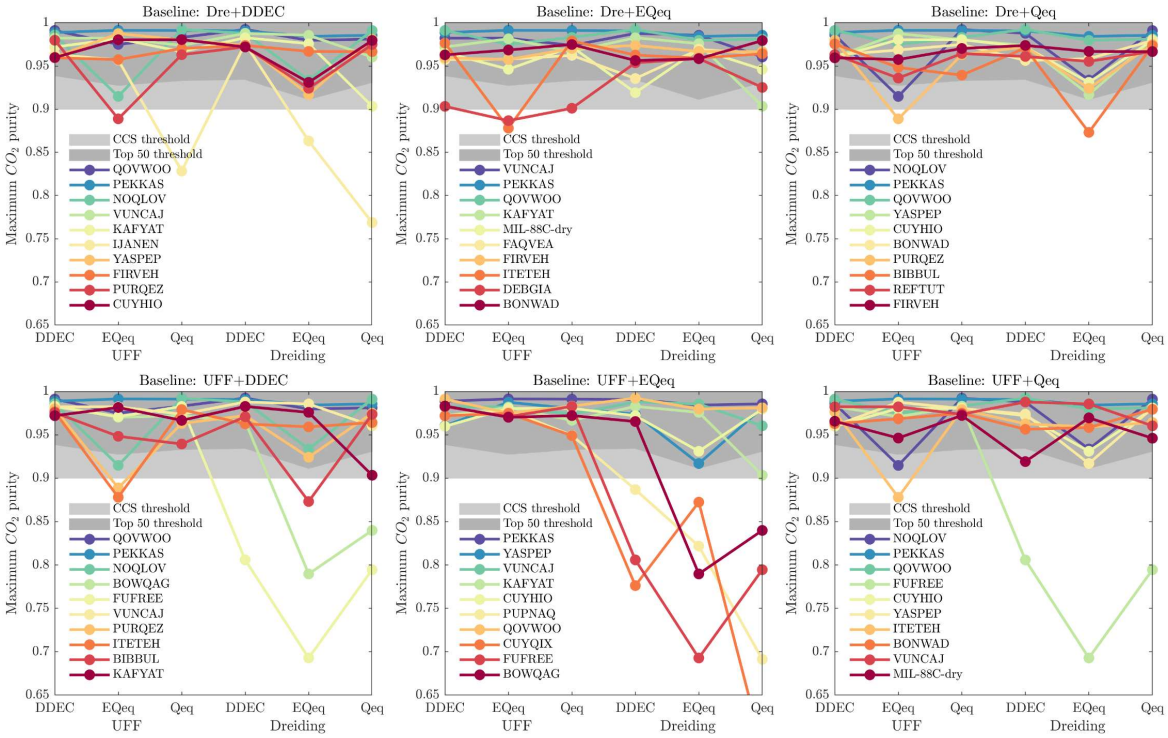

Figure S11. **Visualisation of how the process-level performance values of the top 10 MOFs ranked by different baseline molecular forcefields (title of each subplot) change using other molecular forcefields**. Materials are ranked by the maximum CO_2 purity they can achieve in the modified Skarstrom cycle, subject to the recovery of $CO_2 \geq 0.9$. The dark grey shaded region indicates the process-level performance value which differentiates the top 50 materials in each forcefield. The medium grey shaded region indicates the CCS constraint of CO_2 purity ≥ 0.9. The light grey shaded region indicates the minimum performance value predicted in each forcefield.

Supplementary Note 6: Additional notes and figures for section 3.2 of the main article

S6.1 Comparing the material-level predictions of different forcefields

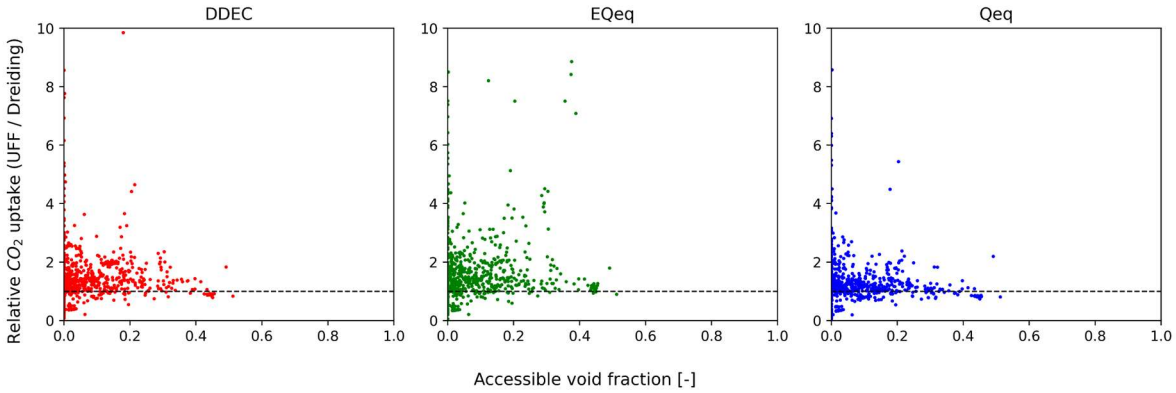

Figure S12. **Relative uptake of CO_2 (UFF predicted uptake / Dreiding predicted uptake) at 298K and 0.1 bar for CRAFTED-u MOFs as a function of the accessible void fraction.**

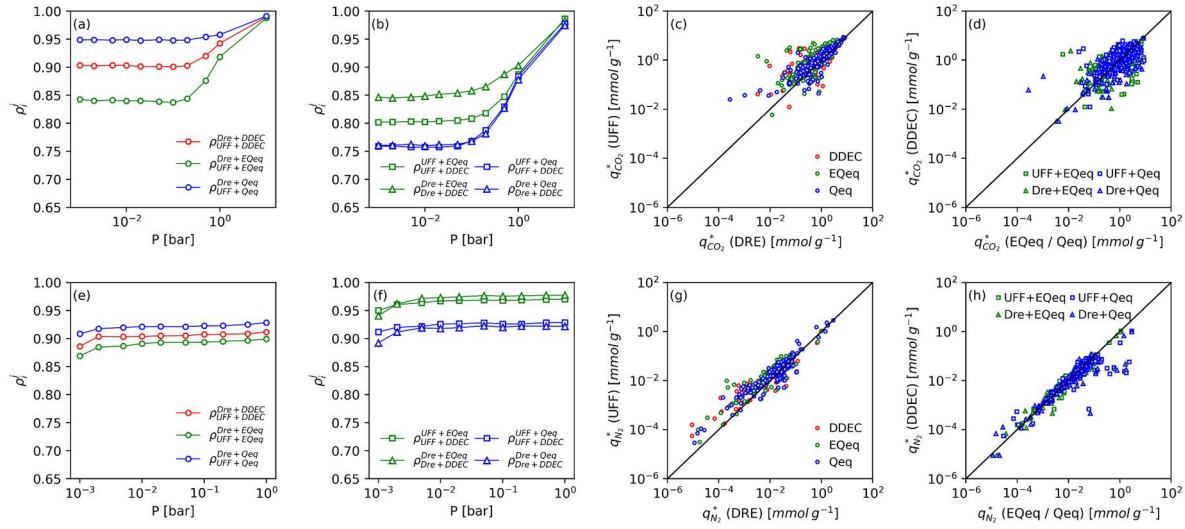

Figure S13. **Correlations and parity plots of simulated uptakes of CO_2 and N_2 for different molecular forcefields.** Subplots (a) and (b) show the Spearman correlations in simulated uptakes of CO_2 for different LJ forcefields and charge schemes, respectively. Subplots (c) and (d) show the parity in simulated CO_2 uptakes at 0.1 bar using different LJ forcefields and charge schemes, respectively. Subplots (e) through to (h) make the same comparisons but for N_2. Note that each point in subplots (c, d) and (g, h) represents a unique material and forcefield. Clusters I and II are highlighted in subplots (c) and (h) respectively to aid in the discussion

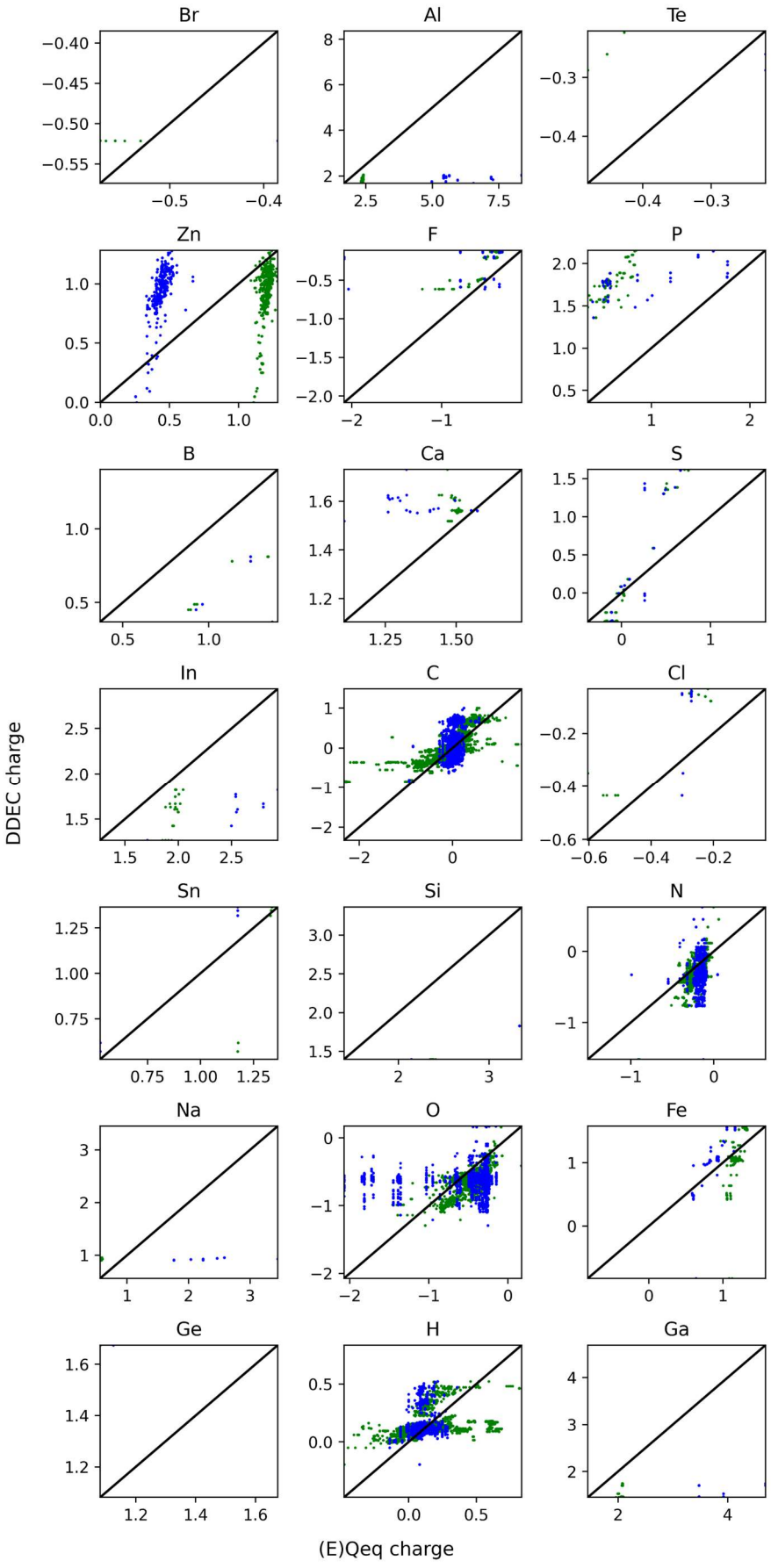

Figure S14. **Comparison of DDEC and EQeq (green points) and DDEC and Qeq (blue points) partial charges for every atom type in the CRAFTED-u dataset**

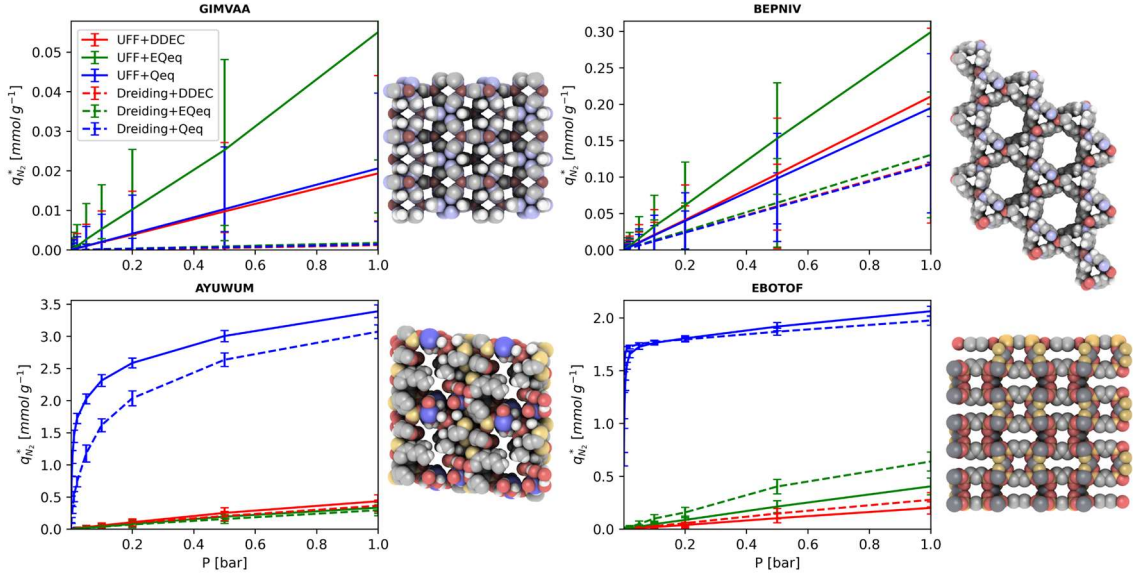

Figure S15. **Distribution of N_2 adsorption isotherms predicted by different molecular forcefields for GIMVAA, BEPNIV, AYUWUM, and EBOTOF.** Error bars are shown at each simulated pressure point. MOF structures are visualised using iRASPA to the right of each isotherm subplot.

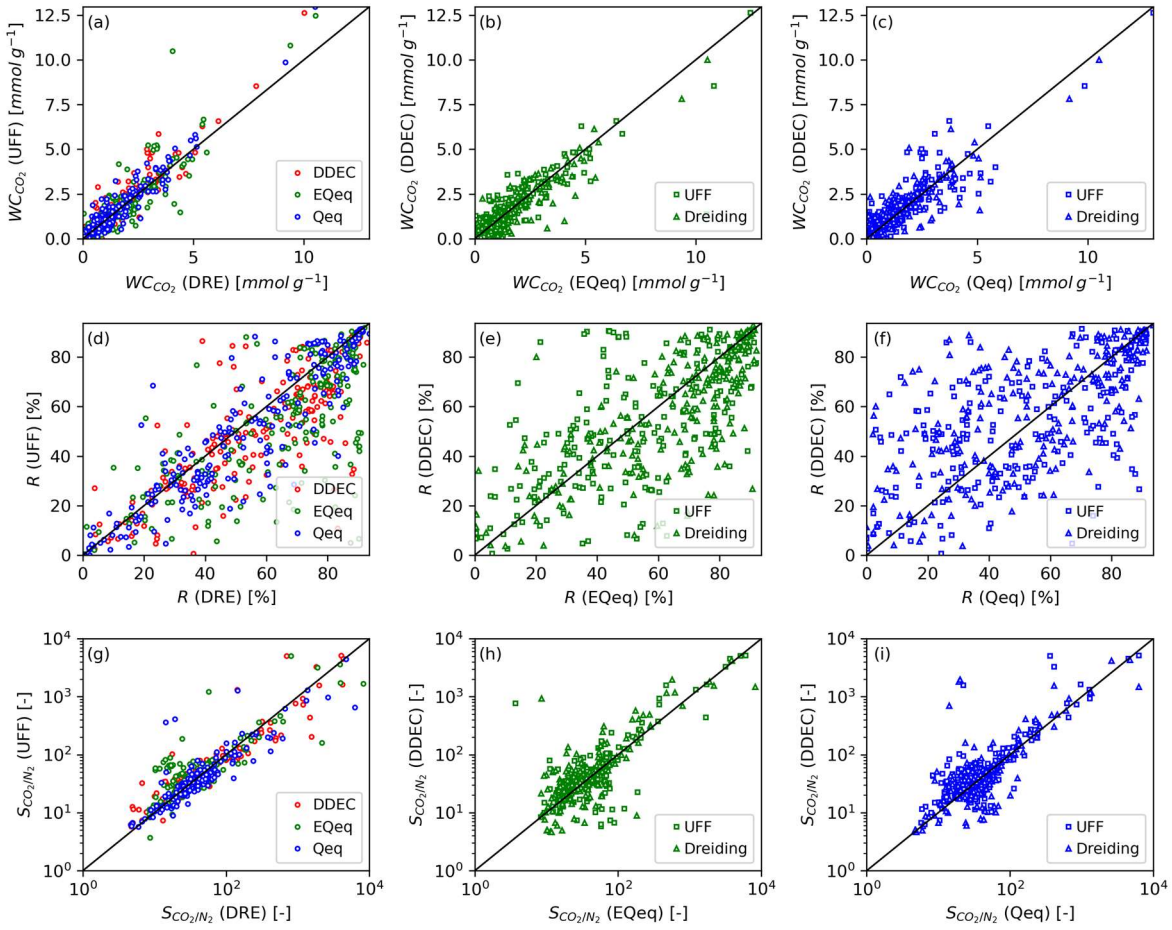

Figure S16. **Parity plots of some common material-level CO_2 capture metrics for different molecular forcefields.** The top row, middle row, and bottom row shows the parity in CO_2 working capacity, percent regenerability, and ideal CO_2 / N_2 selectivity, respectively. The left column is comparing the predictions of the UFF and Dreiding LJ FFs using fixed charges, while the middle and right-most columns are comparing the predictions of DDEC vs EQeq and DDEC vs Qeq using fixed LJ parameters, respectively. Mathematical definitions for each of the material-level KPIs can be found in Table 1 of Farmahini et al. [9]

S6.2 Forcefield Factors for the CRAFTED-u database

The forcefield factor (FFF) or Dokur and Keskin [13] consists of LJ energy parameters, ε_i, and the number of atoms, n_i, for atom type i in the MOF:

$$FFF = \frac{\left|\sum_i^N n_i \times \frac{n_i}{N} \times \frac{\varepsilon_{i,Dreiding}}{k_B} - \sum_i^N n_i \times \frac{n_i}{N} \times \frac{\varepsilon_{i,UFF}}{k_B}\right|}{\sum_i^N n_i \times \frac{n_i}{N} \times \frac{\varepsilon_{i,Dreiding}}{k_B}}$$

Where n_i is the number of atoms i, N is the total number of atoms, and $\frac{\varepsilon_i}{k_B}$ is the energy parameter of atom i. Therefore, the FFF computes how much the potential energy parameter changes when the UFF parameters are used instead of the Dreiding parameters. In Figure S16, we compute the FFF for MOFs in the CRAFTED-u database and show that the FFF is positively correlated with the uncertainty in the adsorption data. Therefore, this factor can help to identify the pool of materials for which the uncertainties between LJ parameter sets are expected to be large. For example, Dokur and Keskin suggest that the performance of MOFs with FFF < 0.1 should be subject to negligible uncertainty, while MOFs with FFF > 0.1 will be non-negligibly impacted by the choice of LJ parameters.

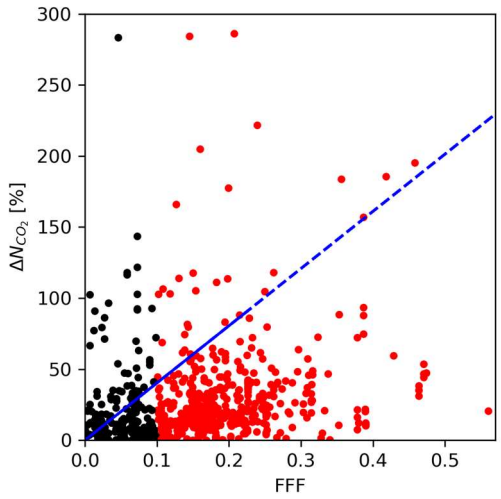

Figure S17. **Relation between the change in the CO_2 uptake and FFF.** The change in CO_2 uptake is calculated as $|N_{CO_2,Dreiding} - N_{CO_2,UFF}| \,/\, N_{CO_2,Dreiding} \times 100$. The dashed blue line shows the line of best fit. Points coloured black are those materials with FFF < 0.1, while those coloured red have FFF > 0.1.

Supplementary Note 7: Additional notes and figures for section 3.3 of the main article

S7.1 Uncertainty heatmap for process-level errors between all pairwise forcefields

In Figure S18, we provide an *uncertainty heatmap* which visualises the average hypervolume error, $\overline{\Delta\xi_i^j}$, between all pairwise FFs i and j. We further partition this map into quadrants Q1 – Q4 to facilitate an easy comparison amongst different molecular modelling approaches. For instance, entries in Q1 and Q3 inform us on the uncertainties that arise from one's choice of LJ FF, while entries in Q2 and Q4 describe the errors between different charge schemes. Darker colours denote higher uncertainties, and vice versa. The underlying $\Delta\xi_i^j$ distributions used to calculate each $\overline{\Delta\xi_i^j}$ value are also provided.

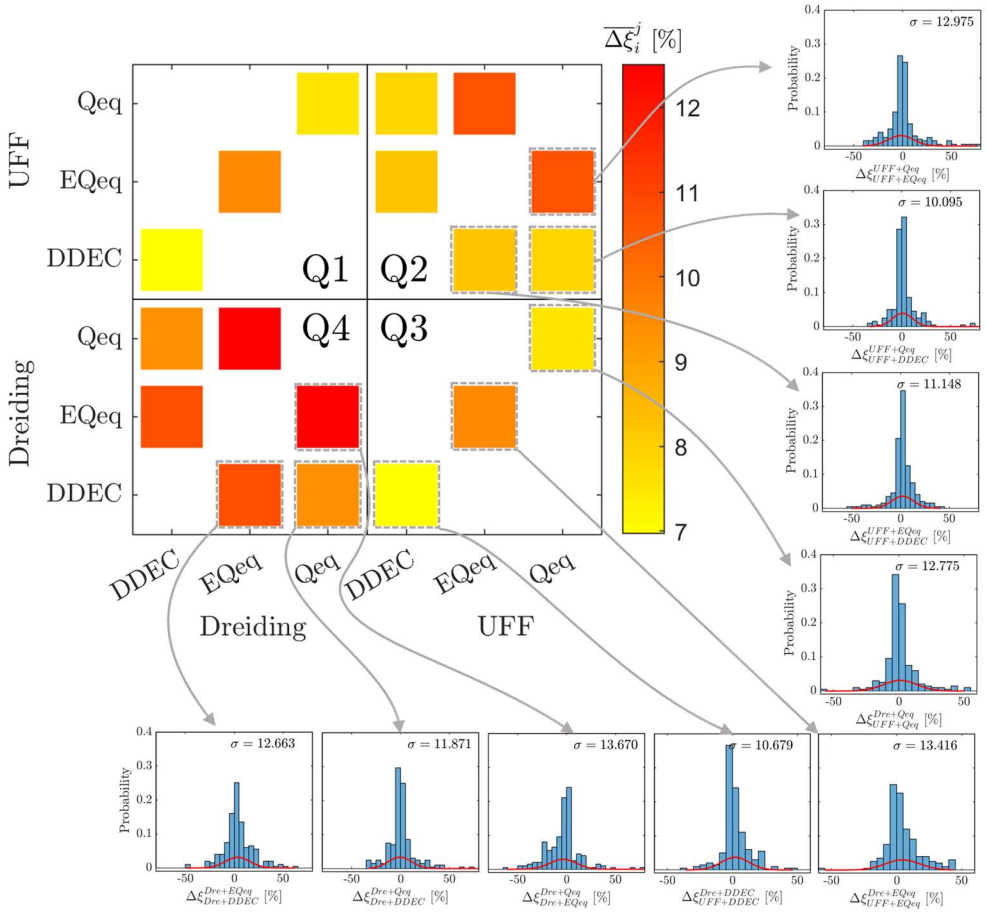

Figure S18. **Uncertainty heatmap for CO_2 purity-recovery optimisations.** Largest subplot visualises the average difference in hypervolumes between all pairwise FF comparisons for the CRAFTED-u dataset. The smaller subplots show the underlying $\Delta\xi_i^j$ distributions which were used in the computation of each $\overline{\Delta\xi_i^j}$ value. A normal distribution is fitted to each histogram (red line), with the standard deviation (σ) from the regression reported in each subplot.

From this figure, we can discern four important trends: (1) for UFF, variations of the charge scheme leads to lower uncertainties in process-level predictions compared to Dreiding, as indicated by the lighter-coloured squares populating Q2; (2) both in the comparisons between different charges (Q2, Q4) and between UFF and Dreiding (Q1, Q3),

the largest deviations arise when the EQeq charge scheme is used; (3) for a pair of FFs i and j, larger errors are typically observed when the same LJ FF but different charge scheme is used, rather than different LJ FF but same charge scheme. That is, the lightest colours fall in quadrants Q1 and Q3; and (4) in situations where the charge scheme is fixed and comparison are made between to LJ FFs, DDEC charges produce the lowest uncertainty overall.

S7.2 Individual material-to-process mappings

Here we briefly demonstrate that individual material behaviours can be mapped to the process-level uncertainty metrics. For convenience, we reproduce some of the figures from the main article here (Figure S19).

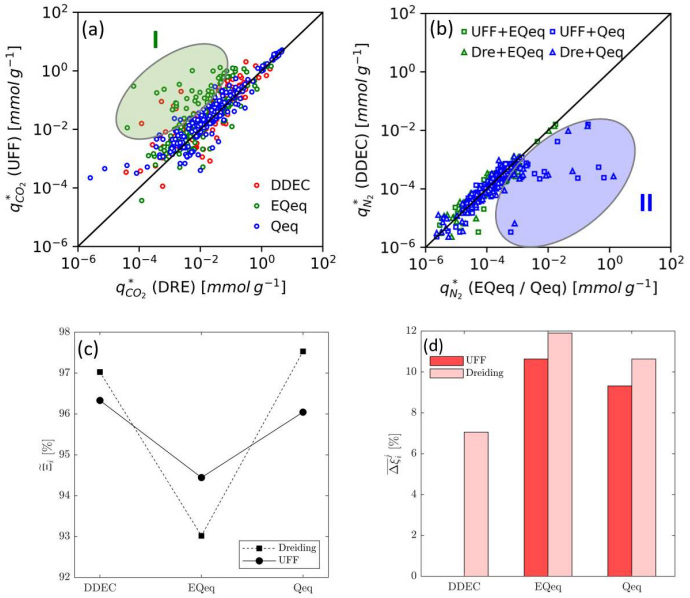

Figure S19. (a, b) Correlations and parity plots of simulated uptakes of CO_2 and N_2 for different molecular forcefields (Figure 8 (c, h) in the main article). (c, d) Process-level uncertainty metrics for CO_2 purity-recovery optimisations (Figure 10 (a, b) in the main article).

Let us revisit the MOFs from cluster II in Figure S19 (b). As a consequence of the high N_2 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 EQeq 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 EQeq *overall*, as evidenced by the uncertainty metrics in Figure S19 (d). 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 EQeq generates $\Delta\xi_i^j$ distributions with greater standard deviations but with fewer statistical outliers. For example, in Table 6 we report the standard deviation and number of statistical outliers for every hypervolume error distribution, $\Delta\xi_i^j$, taking i = UFF+DDEC as the baseline and comparing against every other molecular forcefield. Note that here, statistical outliers are defined as $\Delta\xi_i^j$ values more than 1.5 interquartile ranges above the upper quartile (75%) or below the lower quartile (25%) in each $\Delta\xi_i^j$ distribution.

Table 6. Standard deviations and number of statistical outliers for $\Delta\xi_i^j$ distributions, taking i = UFF+DDEC as the reference.

	$\Delta\xi^{UFF+EQeq}_{UFF+DDEC}$	$\Delta\xi^{UFF+Qeq}_{UFF+DDEC}$	$\Delta\xi^{Dre+DDEC}_{UFF+DDEC}$	$\Delta\xi^{Dre+EQeq}_{UFF+DDEC}$	$\Delta\xi^{Dre+Qeq}_{UFF+DDEC}$
Standard deviation	11.148	10.095	10.679	13.373	12.323
Number of outliers	25	40	24	6	35

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 S19 (a) can be directly linked to the uncertainty results in Figure S19. Taking GIMVAA again as the representative example in this cluster, the UFF+EQeq combination of parameters generates CO_2 isotherms with greater uptakes than the predictions of other FFs (main article, Figure 9, upper left panel). This effect combined with the relatively modest deviations between N_2 isotherms (Figure S15) leads to Pareto fronts which strongly dominate over those generated using other FFs. We therefore see that EQeq in combination with UFF can produce highly favourable process performances such that $\tilde{\Xi}_{UFF+EQeq} > \tilde{\Xi}_{Dreiding+EQeq}$, while the absence of these behaviours using DDEC or Qeq charges leads to the inverse Pareto-dominance relationship observed in Figure S19 (c). 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.

Supplementary Note 8: Additional notes and figures for section 3.4 of the main article

S8.1 List of CoRE 2014 DDEC codes for different performance consistency classes

Table 7. List of CSD codes for CL1 class MOFs

AFOQUI	BENXUP	BONWAD	CUYHIO	FAQVEA	FECWOB01	FECWOB	FIRVEH	HENYUV	KAFYAT
KEXKEF	MIL-88C-dry	NAYZUK	NOQLOV	NURMUJ	OJIWIO	PASMUT	PEKKAS	PURSOK	QOVWOO
QUQPUO	RAJFAL	REFTUT	RURQEB	SUJREV	VUNCAJ	WALCIX	YASPEP		

Table 8. List of CSD codes for CL2 class MOFs

BERFIP	BIBBUL	CATART03	DEBGIA	FAFJIH	GEHSAN	HABRIN	HESVOR	HIWXER	ITETEH
KINKAV	KOMHAW	KUNWEW	OSOMIT	PARHAS	PURQEZ	REYCOP	RUTBUE	WALBOC01	WALBOC02
WALBOC	XOVVOU								

Table 9. List of CSD codes for CL3 class MOFs

ABUWOJ	ACUBAB	AMUCOB	APEBED	ASACON	ATOXEN	AXOCEW	AYUWUM	BENXOJ	BEPLUF
BEPLUF	BEPMAM	BEPNIV	BEPNOB	BEPPIX	BEPPOD	BOWQAG	CAHSOU	CAXVII	CAXVOO
CAXVUU	CAXWAB	CAXWEF	CAXWIJ	CISMAT01	COF-1	CUNXIS10	CUYQIX	DEBNUU	DEPXAY
EBIHII	EBOTOF	ECASAD	ECUTON	ESEVIH	EXEWAG	FAGREM	FEHCOM	FIPWOS	FIPXEJ
FIQYOU	FIVYUF	FUFREE	GELJAJ	GIHBII	GIMVAA	GULPIN	GUQTOB	HAFQUC	HAJLEK
HAJLEK	HAJLIO	HAJLOU	HAJLUA	HAJMAH	HAJMEL	HELCUY	HEZKIH	HIWXER01	HUTZAX01
HUTZEB	IBIDAA	IFENOY	IFEPUG	IFEQAN	IFEQIV	IGOCOX	IJANAJ	IJANEN	IJEXUR
IMACZN01	IYUJAO	JITPOS	KENJEU	KEXKAB	KEXPOU	KINHAS	KOMGUP	LACGAZ	LACJAC
LADHEF	LAZJAZ	LEGGOU	LEMNOH	LIBJAJ	LIFWOO	LIKFOB	LIRNEH	LIWXIZ	LOLQIM
LUPHIN	LUPTAS	MAFZOK	MEFYUT	MIDRAT	MIKJAR	MOYNEU	MULQOA	NARTUX	NIKYOW01
NIKYOW01	NIKYOW	NIKZAJ01	NIKZAJ	NIKZEN01	NIKZEN	NIRDOI	NOFHUM	NUCREJ	OSAVEK
OSAVUA	OYEBIE	OYODEM	PARHEW	PEGBAG	PEGBEK01	PEGBEK02	PEGBUA	PEGCAH	PEGCEL
PENNUT	PUPNAQ	PUQHIT	PUWCUF	QATHIE	QIFLUO	QUFFED	RACZIG	RAXDAX	REDROI
REDYEG	REYBAA	RIPNIP	RIYDEJ	RUFMUA01	RUFMUA	SADMOB	SENWAL	SENWEP	SUJQOE
SUJQOE	TADCAE	TEBSAW	TECVEE	TEVZOK	TISKOW	UBACOR	UJAQUS	UMEXUF	VEHKAW
VONBIK	WALCET	WAYMEQ	WEBREC	XEXLUJ	XIDBUJ	XILDUT	XIMPOA	XINWUO	XISVUR
XOHLEM01	XOHLEM	XOKHAH	XOVVEK	XOVVIO	YARKAF	YAXBOP	YAZFOW	YECQUS02	YIKLUZ
YILJAG	YUBFOR	YUKBIP01	YUSGID	ZARZAV	ZEDROR	ZIF-5	ZNGLUD01	ZNGLUD	Zn-DOBDC

Table 10. List of CSD codes for CL4 class MOFs

ACOLIP	AFITIT	AHINIP	AMILUE	AMIMEP	ANUGUM	ATAFIK	ATOWIQ	ATOWOW	AVIMOI
AXUBAW	AZIXUD	BARZAW	BARZUR	BEPMEQ	BEPMIU	BEPNUH	BEPPAP	BERGAI	BICDAU
BIYTEJ	BOLPEY	BUKMUQ	BUVXOG	BUVYEX	BUVYIB	CADPII	CANRUG	CAWVIH	CEGWER
CELZOK	COF-300	COF-5	COGWEB	COMDOY	CUTKOS	CUYVOI	DAXGUG	DAXHIV01	DAXHIV
DAXHUH01	DAXHUH	DAXKOE	DAXNOG	DEJCEB	DEYVUA	DIDDOK	DOGZIJ	DONNAW01	DORDUK
DOTTUC	DUBWON	ECAHAT	ECIWUJ	EDADIX	EDUSIF	EGEJIK	EGELUY01	EMITUQ	EMIVAY
EZUCIM	FANWOI	FASQUN	FAYKOH	FECKAA	FECZAQ	FEFCUQ	FEFDAX	FEFDEB	FERHAN
FIJCUX01	FIJCUX	FIPWUY	FUDQIF	FUNBAS	FUNBEW	FUNBIA	FUNBOG	FUNCAT	FUNCEX
FUTDII	FUTKEL	FUWXOL	GACPUX	GACQAE	GALCAZ	GAXVUY	GAYFOD	GEDLIM	GEGSIW
GERNOI	GIDKOU	GINLIA	GITWIQ	GIYTIS	GIZJOP	GOGWAB	GOMREG	GUYLOC	GUYLUI01
GUYLUI	GUYMAP	HAFQOW	HAFVOB	HAMJOW	HAXJEX	HEBTEP	HEBZAR	HEBZEV	HEDCEA
HEGJUZ	HELDEJ	HICVOG	HIDMEO	HIFGIO	HIFGOU	HIFTOG01	HIFTOG02	HIFTOG	HIFVOI
HIFVUO	HIFWAV	HIHGOW	HIHJUF	HIHNUJ	HIMSAY	HITXUE	HOMZEP	HOWPUF	IBICAZ
IBICED	IBICIH	IBICON	IBICUT	IBOTAU	IBUYAH	ICAGAU	ICALOP	IFENOY01	IFENOY02
ILOJEZ	IRMOF-1	IRMOF-6	IRMOF-7	ISEQIH	ISOHEE	ISOJOQ	ISUCUV	IXURAV	IXUREZ
IXURID	IYIHUU	JASNAT	JASNEX	JIVFUQ	KAFTUI	KAMZUV	KANCIN	KANDEK	KANDIO
KAVROQ	KAXQIL	KAXQOR01	KAXQOR	KAYBIX	KAYBUJ	KAYCIY	KEDQAN	KEFBEE	KEFBOO
KEWZOD	KIDDOS	KIGCEK	KIXXOG	KIYMIQ	LABGAY	LAWGEW	LAWGIA	LAWGOG	LAWGUM
LEGGEK	LEGGIO	LETRIN	LETROT	LIDZUV	LOJFUL	LOQLEJ	LOQLIN	LOWPUI	LOWQIX
LURGEL	LUVTEC	MAGBAZ	MAKXAZ	MAPFIS	MAPFOY	MATTOR	MATTUX	MATVAF	MATVEJ
MATVIN	MAZSUD	MECWEX	MEJQEZ	MEJQID	MEJQOJ	MIBQAR16	MIBQAR18	MIBQAR	MIFKUJ
MIL-88A-dry	MIL-88A-open	MIL-88C-open	MIMVEJ	MIZJUB	MOF-1-Dybtsev-dmol	MOF-69	MUVGOA	MUVGUG	MUVHAN
MUVHER	MUWQEB	MUZHUL	NAKLIW	NAQLAT	NAZBAT	NEHNEV01	NEHNEV	NEHZIK	NHBZZN10
NIHBEM01	NIHBEM02	NIHBEM	NIHBIQ01	NIHBIQ02	NIHBIQ	NIKYOW02	NIKYUC01	NIKYUC02	NIKYUC
NIKZAJ02	NIKZEN02	NISCOJ	NOCKUM	OBUFEX	OBUFIB	OCIZIL	ODIXEG	ODODES	ODODIW
ODUNEH	OFERUN02	OFERUN04	OFUYUL	OGAGAF	OGAGEJ	OGAGIN	OHOLIH	OSAWEL	OSAXAI
OVICUS	OWAVOY	PAMVEG	PAPHEV	PEJFOA	PEMRIK	PENZUE	PEPBAO	PEQHUQ	PEVPUD
PEVQAK	PEVQEO	PEVQIS	PEVQOY	PEYPIU	PIDMEV01	PIDMEV	PIFPIE	PIHNUQ	PIJJOI
PIXHOV	PIXPET	PORVUO	PUQXUV	PUQYAC	PURJES	PURJIW	PUSDIR	PUSDOX	QAQSAD
QARCET	QAVWAN	QEDZEG	QIFLIC	QIFLOI	QOKCID	QOMDUS	QOMFAA	QUDJOP	QUPHUF
QUQFIS	QUQPOI	RAHBUZ	RAPYOY	RAPYUE	RAQVEM	RATCUL01	RATCUL02	RATCUL03	RATDAS02
RAXCOK	REGQUR	REGREC	RELLAW	RETBEZ	REYPIW	RIFDUG01	RUMWUS	RUPZIM	RUVMAX
SABVOH	SADMUH	SAHYOQ03	SAHYOQ04	SAHYOQ05	SAKRAZ	SAKRED	SARBOE	SARNOQ	SEFBOV
SENWIT	SENWOZ	SETQOZ	SETQUF	SULKIU	TACPAP	TAPHOJ	TASXIW	TATPOV	TECTUS
TEDGOA	TEDHAN	TEDLUK	TEDSIG	TENLIJ	TEPGUS	TEQPAI	TESGOO	TESGUU	TESHAB
TEVZEA	TEVZIE	TEVZUQ	TIGDOD	TIRLIQ	TONBII	TONQOD	TUDJOS	UBACUX	UBUMAH
UDANAQ	UKUBUY	ULUVUT	UNIGEE	UTEXAT	UVARIT	UVAROZ	UVARUF	UVASAM	UVEXAV
UVUPOR	UVUQAE	UWELIS	UWUTIQ	UXEHIP	UXUZAP	VACFOV01	VACFOV	VACFUB01	VACFUB
VAGTUU	VAHSIH	VAZTOG	VEJYIT01	VEJYIT	VEJYUF01	VEJYUF02	VEMMAC	VEXVAW	VICDOC
VICYUD	VIZRIH	VONBUW	VUSJUP	VUSKAW	VUSKEA	WAFKAQ	WEBKOF	WECJAR	WEMNAF
WENDIE	WIJDAV	WIYFAM	WOGHEG	WOJJOV	WONZOP	WONZUV	WUCRUH	XADCOW	XADGAM
XAJQAC	XAMHEA	XAXQEU	XECBUD	XEDPIH	XEDPON	XENKIM	XENMIO	XENMOU	XEQGEH
XEXLIX	XEXLOD	XEXMEU	XINFUW	XOQXEH	XUGSEY	XUNJEW	XUVHEB	XUWVEQ	XUYXAR
YARFII	YARGAB	YARKEJ	YARSAN	YEGKOM	YEKXET	YEZFIU	ZEDHIB	ZERQOE	ZESFUY
ZIDDIB	ZIF-1	ZIF-2	ZIF-3	ZIF-4	ZIF-6	ZIF-77	ZIF-7	ZIF-8	ZIF-90

ZnBDC	ZnBDCdabco	ZnHBDC	ZnMOF-74	ASALIP	BEGBOG	BEKSAM	CUNXIS	EGELUY	EHALOP
FABFOF	GUXLIU	HAKWUM	HIWLOO	HUTTIA	ISIMAZ	JAVTAC	LAJDUV01	LAJDUV02	LAJDUV03
LAJDUV	LIHRIE	MOGVAG	PEPXIS	POJTOY01	POJTUE	QONQEQ	REYNUG	SABVOH01	SABVUN
SICDAL	WAYMIU	WAYMOA	WERKUZ	WOWGEU	YUWHOO	EGELUY	EHALOP	LADJUV01	QADBII
QONQEQ	WAYMOA	MECLEL	SABVOH01	SABVUN	WAYMIU	WIHCUM	XEDLUN	XEDLUV	

S8.2 Isotherms of CL1 class MOFs

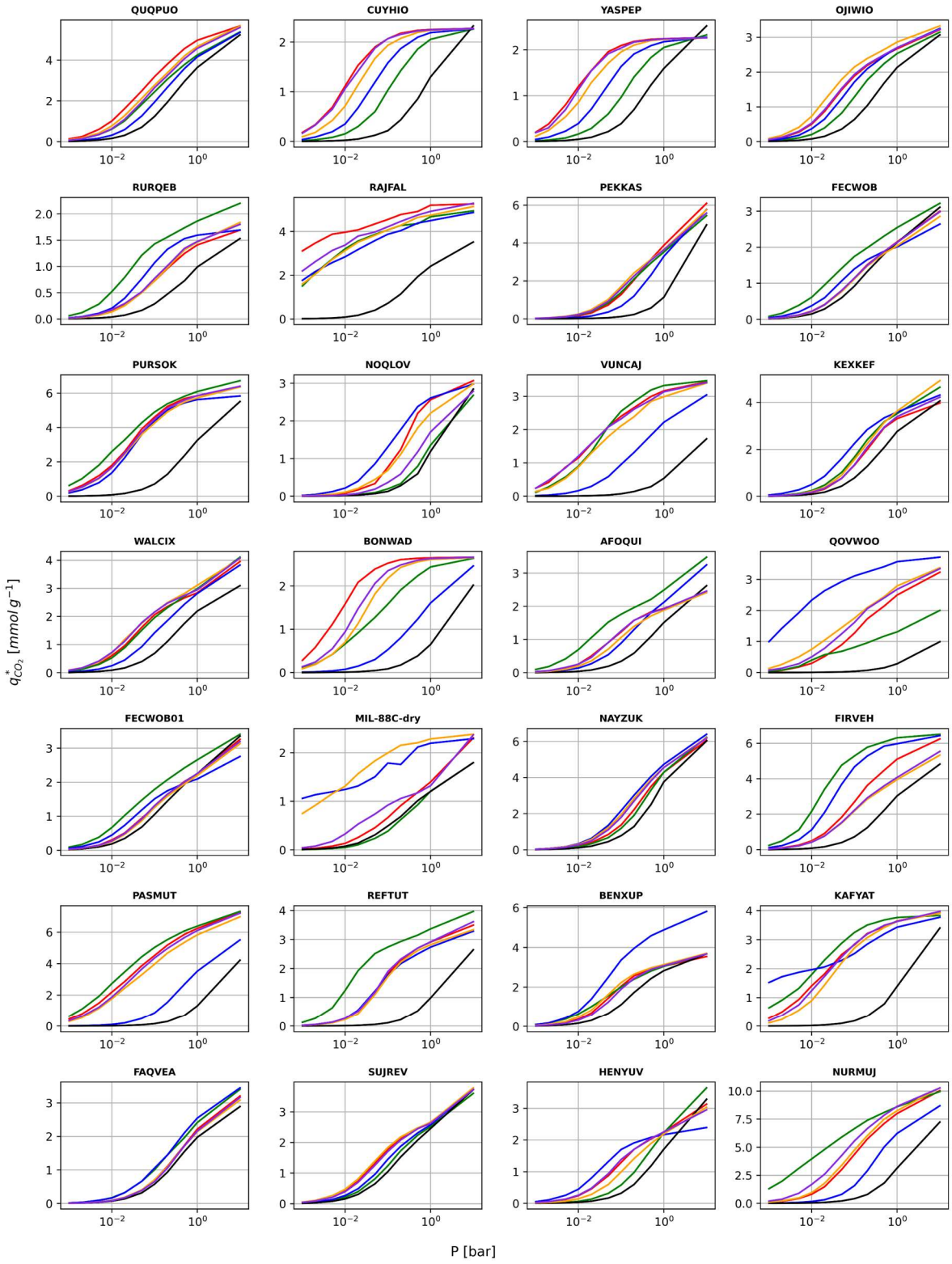

Figure S20. **Distribution of CO_2 adsorption behaviours in CL1 class MOFs at 298K arising from different charge schemes and fixed UFF forcefield.** Red = DDEC, green = EQeq, blue = Qeq, purple = MPNN, orange = PACMOF, black = neutral. FECWOB, FIRVEH, MIL-88C-dry, NOQLOV, RURQEB, SUJREV, and WALCIX were not used in the training of PACMOF.

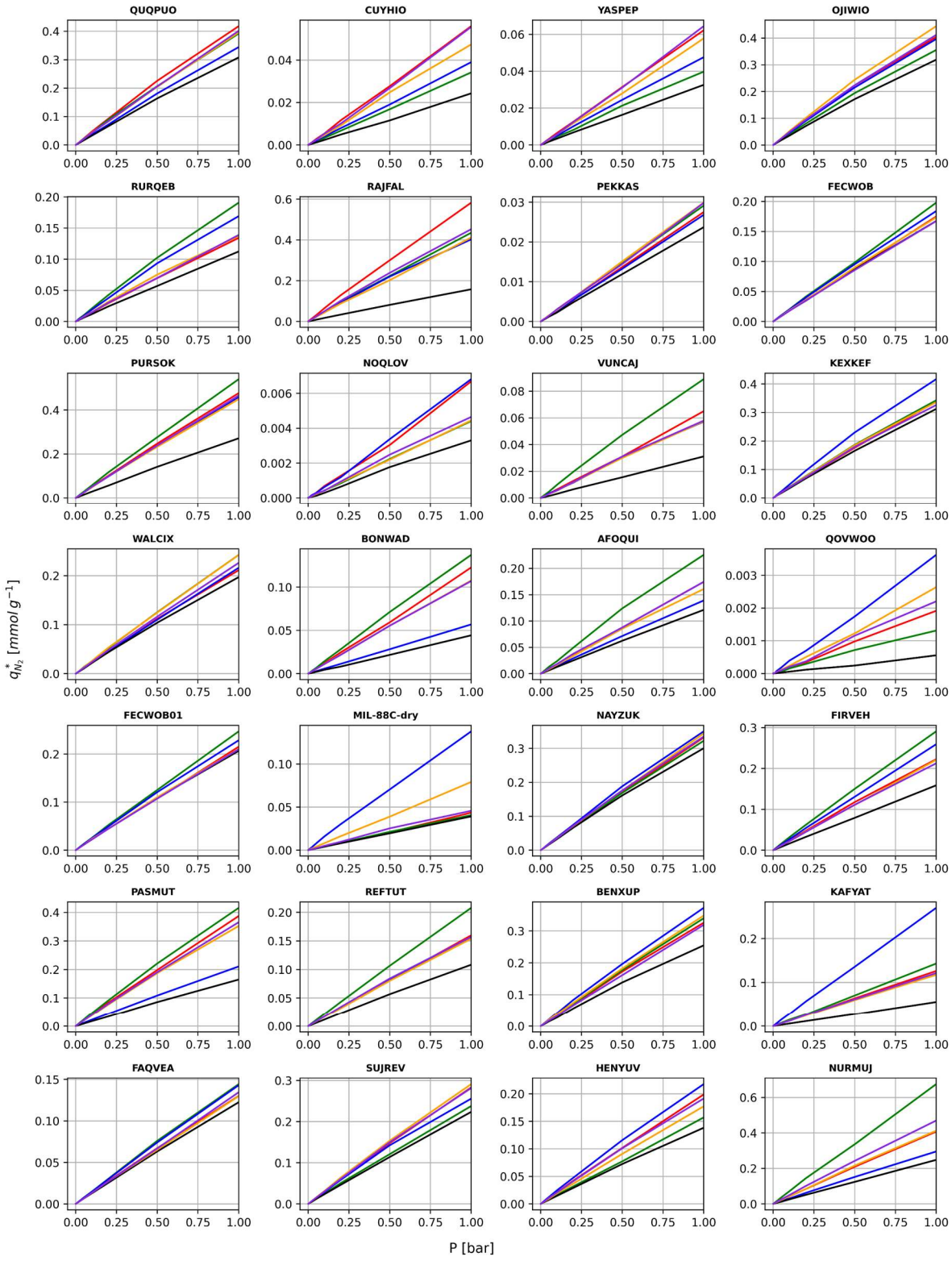

Figure S21. **Distribution of N_2 adsorption behaviours in CL1-class MOFs at 298K arising from different charge schemes and fixed UFF forcefield.** Red = DDEC, green = EQeq, blue = Qeq, purple = MPNN, orange = PACMOF, black = neutral. FECWOB, FIRVEH, MIL-88C-dry, NOQLOV, RURQEB, SUJREV, and WALCIX were not used in the training of PACMOF.

S8.3 Average energy-productivity process-level uncertainty between UFF and Dreiding

We identify 77 MOFs from the CRAFTED-u database which meet the CCS constraints (CO_2 purity and recovery ≥ 0.9) for both UFF+DDEC and Dreiding+DDEC forcefields. We optimise their energy-productivity performances and calculate the hypervolume errors between Pareto fronts generated using both UFF+DDEC and Dreiding+DDEC. On average, predictions between LJ FFs deviate by approximately 30% (Figure S22). The Spearman correlation between the rankings obtained using both FFs is 0.553.

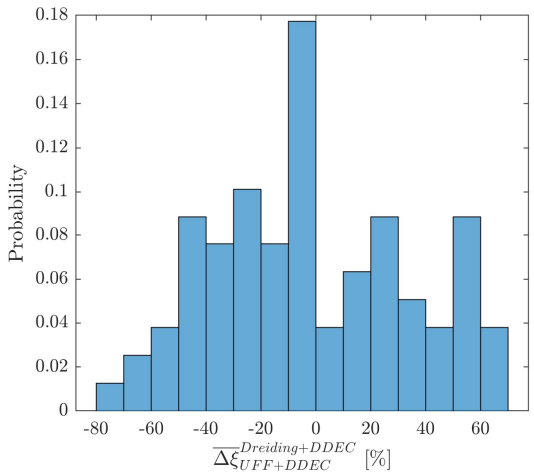

Figure S22. **Hypervolume error distribution between energy-productivity Pareto predictions using UFF+DDEC and Dreiding+DDEC forcefields.** All 77 MOFs which meet the CCS constraints using either UFF or Dreiding in combination with DDEC charges used to calculate the average hypervolume error of 29.6%.

Supplementary note 9: Comment on the application of ML-based charges to CL1 class MOFs

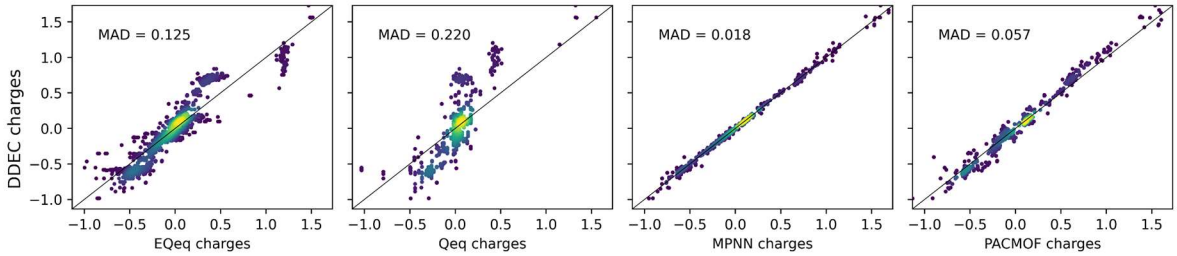

Figure S23. **Parity plots of DDEC charges vs charges predicted by charge equilibration and machine learning methods for atoms in CL1 class materials.** Colour indicates the density of points (number of atoms) that fall in that bin. Diagonal line shows $y = x$.

Parity plots comparing the charges assigned to framework atoms by the DDEC charge scheme versus the EQeq, Qeq, MPNN, and PACMOF charge schemes for CL1 class MOFs are provided in Figure S23. Here we comment on the biases present in the machine learning (ML) charge assignment models and how, in particular, these biases are not expected to significantly impact the uncertainty results discussed in section 3.4 of the main article.

The DDEC charges of all CL1 class MOFs were used in the training of the MPNN model. Despite the significant data leakage (that is, when data from the training set is used in the testing set) of MPNN charges, the mean absolute deviation (MAD) of 0.018 in Figure S23 is not dissimilar to the MAD reported on the test dataset in the original publication (0.025) [14]. We therefore do not believe that our results in section 3.4 of the main article are a significant departure from the uncertainties that would be observed for materials absent in the training of the MPNN model. To support this argument, consider the materials FECWOB, FIRVEH, MIL-88C-dry, NOQLOV, RURQEB, SUJREV, and WALCIX. These MOFs were not used in the training of the PACMOF model, and therefore the adsorption behaviours and process-level results are purely predictive for these materials. We see from Figure S19 and

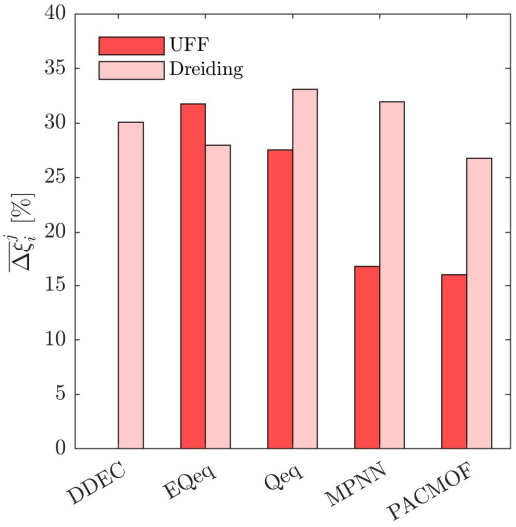

Figure S24. **Process-level uncertainty metrics for energy penalty-productivity optimisations of MOFs not used in the training of the PACMOF model.** Average hypervolume error, $\overline{\Delta\xi}_i^j$, between i = UFF+DDEC and other molecular forcefields ($j \neq$ UFF+DDEC). Higher values of $\overline{\Delta\xi}_i^j$ indicate a larger average uncertainty with the UFF+DDEC baseline. FECWOB, FIRVEH, MIL-88C-dry, NOQLOV, RURQEB, SUJREV, and WALCIX were not used in the training of PACMOF.

Figure S21 that, out of all these materials, the agreement between PACMOF and DDEC adsorption isotherms are only worse than the (E)Qeq adsorption isotherms in MIL-88C-dry. Furthermore, if we calculate the hypervolume errors between the energy-productivity Pareto fronts that are generated using different molecular forcefields for only these 8 materials, we see that the process-level uncertainties in Figure S24 are qualitatively similar to those reported in Figure 11 in the main article. In particular, the uncertainties between UFF and Dreiding using fixed DDEC charges are larger than the uncertainties between DDEC and PACMOF using fixed UFF parameters. This means that the error that is introduced by PACMOF in a purely predictive capacity is still substantially less than the errors introduced by using charge equilibration schemes or by one's choice in LJ parameters.

Supplementary Note 10: Comparing the performance of CALF-20 to CL1 class MOFs

In Figure S25, the energy-productivity process-level performance of CALF-20 [15] is shown in red, while the performances of other CL1 class MOFs (using the UFF+DDEC molecular FF) are shown in black. Dual-site Langmuir model parameters (Table 11) for CALF-20 were fitted to the experimental isotherm data reported in [15]. Out of 29 materials (28 MOFs in CL1 class plus CALF-20), CALF-20 is ranked 20th in terms of its energy-productivity performance.

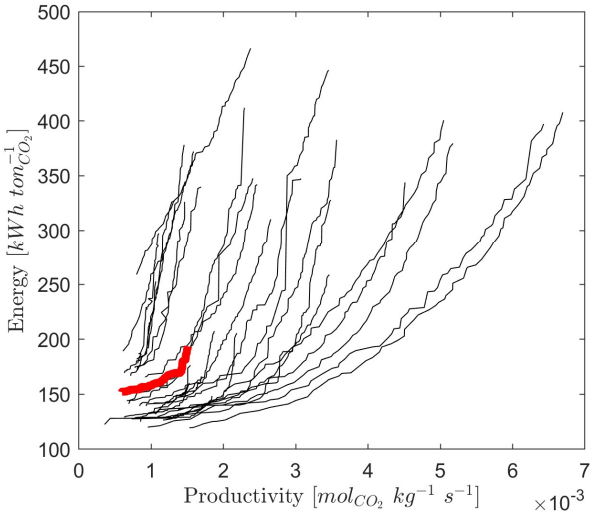

Figure S25. **Energy-productivity performance curves of CL1 class MOFs (black) and CALF-20 (red).**

Table 11. Summary of DSL model parameters used to model CALF-20

DSL parameter	Units	Value
$q_{s,1}^{CO_2}$	$[mmol\ g^{-1}]$	2.6127
$q_{s,2}^{CO_2}$	$[mmol\ g^{-1}]$	3.2760
$b_{0,1}^{CO_2}$	$[bar^{-1}]$	1.1045x10^{-5}
$b_{0,2}^{CO_2}$	$[bar^{-1}]$	1.0756x10^{-7}
$\Delta H_{CO_2,1}^{ads}$	$[J\ mol^{-1}]$	-37022
$\Delta H_{CO_2,2}^{ads}$	$[J\ mol^{-1}]$	-37022
$q_{s,1}^{N_2}$	$[mmol\ g^{-1}]$	2.6127
$q_{s,2}^{N_2}$	$[mmol\ g^{-1}]$	3.2760
$b_{0,1}^{N_2}$	$[bar^{-1}]$	9.6559x10^{-6}
$b_{0,2}^{N_2}$	$[bar^{-1}]$	9.6559x10^{-6}
$\Delta H_{N_2,1}^{ads}$	$[J\ mol^{-1}]$	-21046
$\Delta H_{N_2,2}^{ads}$	$[J\ mol^{-1}]$	-21046

References

[1] D. Dubbeldam, S. Calero, D.E. Ellis, R.Q. Snurr, D. Dubbeldam, S. Calero, D.E. Ellis, R.Q. Snurr, RASPA : molecular simulation software for adsorption and diffusion in flexible nanoporous materials, Mol. Simul. 42 (2016) 81–101. https://doi.org/10.1080/08927022.2015.1010082.

[2] K.P. White, An Effective Truncation Heuristic for Bias Reduction in Simulation Output, Simulation. 69 (1997) 323–334.

[3] C. Cleeton, A.H. Farmahini, L. Sarkisov, Performance-based ranking of porous materials for PSA carbon capture under the uncertainty of experimental data, Chem. Eng. J. 437 (2022) 135395. https://doi.org/10.1016/j.cej.2022.135395.

[4] D. Yancy-Caballero, K.T. Leperi, B.J. Bucior, R.K. Richardson, T. Islamoglu, O.K. Farha, F. You, R.Q. Snurr, Process-level modelling and optimization to evaluate metal–organic frameworks for post-combustion capture of CO 2 , Mol. Syst. Des. Eng. 5 (2020) 1205–1218. https://doi.org/10.1039/d0me00060d.

[5] L.F. Shampine, M.W. Reichelt, The MATLAB ode suite, 1997. https://doi.org/10.1137/S1064827594276424.

[6] R. Kumar, V.G. Fox, D.G. Hartzog, R.E. Larson, Y.C. Chen, P.A. Houghton, T. Naheiri, A versatile process simulator for adsorptive separations, Chem. Eng. Sci. 49 (1994) 3115–3125. https://doi.org/10.1016/0009-2509(94)E0085-5.

[7] A. Nalaparaju, M. Khurana, S. Farooq, I.A. Karimi, J.W. Jiang, CO2 capture in cation-exchanged metal-organic frameworks: Holistic modeling from molecular simulation to process optimization, Chem. Eng. Sci. 124 (2015) 70–78. https://doi.org/10.1016/j.ces.2014.09.054.

[8] H. Jiang, A.D. Ebner, J.A. Ritter, Importance of Incorporating a Vacuum Pump Performance Curve in Dynamic Adsorption Process Simulation, Ind. Eng. Chem. Res. 59 (2020) 856–873. https://doi.org/10.1021/acs.iecr.9b04929.

[9] A.H. Farmahini, S. Krishnamurthy, D. Friedrich, S. Brandani, L. Sarkisov, Performance-based screening of porous materials for carbon capture, Chem. Rev. (2021). https://doi.org/10.1021/acs.chemrev.0c01266.

[10] S.M. Moosavi, B.Á. Novotny, D. Ongari, E. Moubarak, M. Asgari, Ö. Kadioglu, C. Charalambous, A. Ortega-Guerrero, A.H. Farmahini, L. Sarkisov, S. Garcia, F. Noé, B. Smit, A data-science approach to predict the heat capacity of nanoporous materials, Nat. Mater. (2022) 1–28. https://doi.org/10.1038/s41563-022-01374-3.

[11] T.F. Willems, C.H. Rycroft, M. Kazi, J.C. Meza, M. Haranczyk, Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials, Microporous Mesoporous Mater. 149 (2012) 134–141. https://doi.org/10.1016/j.micromeso.2011.08.020.

[12] J. Beck, D. Friedrich, S. Brandani, E.S. Fraga, Multi-objective optimisation using surrogate models for the design of VPSA systems, Comput. Chem. Eng. 82 (2015) 318–329. https://doi.org/10.1016/j.compchemeng.2015.07.009.

[13] D. Dokur, S. Keskin, Effects of Force Field Selection on the Computational Ranking of MOFs for CO2 Separations, Ind. Eng. Chem. Res. 57 (2018) 2298–2309. https://doi.org/10.1021/acs.iecr.7b04792.

[14] A. Raza, A. Sturluson, C.M. Simon, X. Fern, Message Passing Neural Networks for

Partial Charge Assignment to Metal-Organic Frameworks, J. Phys. Chem. C. 124 (2020) 19070–19082. https://doi.org/10.1021/acs.jpcc.0c04903.

[15] J.-B. Lin, T.T.T. Nguyen, R. Vaidhyanathan, J. Burner, J.M. Taylor, H. Durekova, F. Akhtar, R.K. Mah, O. Ghaffari-Nik, S. Marx, N. Fylstra, S.S. Iremonger, K.W. Dawson, S. Partha, P. Hovington, A. Rajendran, T.K. Woo, G.K.H. Shimizua, A scalable metal-organic framework as a durable physisorbent for carbon dioxide capture, Science (80-.). 374 (2021) 1464–1469.