A Robust Framework for Generating Adsorption Isotherms to Screen Materials for Carbon Capture

Elias Moubarak,† Seyed Mohamad Moosavi,†‡ Charitheas Charalambous,¶
Susana Garcia,¶ and Berend Smit*,†

†Laboratory of Molecular Simulation, Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), Rue de l’Industrie 17, CH-1951 Sion, Valais, Switzerland
‡Department of Chemical Engineering & Applied Chemistry, University of Toronto, Ontario M5S 3E5, Canada
¶The Research Centre for Carbon Solutions (RCCS), School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AS Edinburgh, United Kingdom

E-mail: berend.smit@epfl.ch

Abstract

To rank the performance of materials for a given carbon capture process, we rely on pure component isotherms from which we predict the mixture isotherms. For screening a large number of materials we also increasingly rely on isotherms predicted from molecular simulations. In particular, for such screening studies, it is important that the procedures to generate the data are accurate, reliable, and robust. In this work, we develop an efficient and automated workflow for a meticulous sampling of pure component isotherms. The workflow was tested on a set of metal-organic frameworks
(MOFs) and proved to be reliable given different guest molecules. We show that the coupling of our workflow with the Clausius-Clapeyron relation saves CPU time, yet enables us to accurately predict pure component isotherms at the temperatures of interest, starting from a reference isotherm at a given temperature.

We also show that one can accurately predict the CO\textsubscript{2} and N\textsubscript{2} mixture isotherms using Ideal Adsorbed Solution Theory (IAST). In particular, we show that IAST is a more reliable numerical tool to predict binary adsorption uptakes for a range of pressures, temperatures, and compositions, as it does not rely on the fitting of experimental data, which typically needs to be done with analytical models such as dual-site Langmuir (DSL). This makes IAST a more suitable and general technique to bridge the gap between adsorption (raw) data and process modelling. To demonstrate this point, we show that the ranking of materials, for a standard 3-step Temperature Swing Adsorption (TSA) process, can be significantly different depending on the thermodynamic method used to predict binary adsorption data. We show that for the design of processes that capture CO\textsubscript{2} from low concentration (0.4%) streams, the commonly used methodology to predict mixture isotherms incorrectly assigns up to 33% of the materials as top-performing.

Introduction

Carbon dioxide, considered one of the main greenhouse gasses, has a major impact on climate change. Its atmospheric concentration is already above 400 ppm and is predicted to hit alarming levels in the 2050s\textsuperscript{1} if the current situation is not addressed properly. Emission sources are numerous and contribute differently to the problem. They range from deforestation and cement manufacturing to the burning of fossil fuels such as natural gas, petroleum, and coal. Consequently, sectors relying heavily on fossil fuels, such as power generation, industrial, and transportation sectors, are the main sources of CO\textsubscript{2} emissions worldwide.\textsuperscript{2} Current carbon capture and storage technologies have the potential to mitigate climate
change. However, they are energy intensive. Chemical absorption using amine solutions to recover CO₂ from flue gas streams is well developed and deployed at a commercial scale. Also, polymeric membranes are being used commercially to separate CO₂ from syngas. Reaching the full potential within the time limit imposed by the fact that CO₂ emissions are still increasing has motivated research to capture CO₂ directly from atmospheric air. Adsorption-based technologies are promising candidates that address these challenges; to efficiently capture CO₂ from a wide range of sources.

To accelerate the pace of development and provide cost-effective solutions for carbon capture applications, the materials and process engineering research communities have shifted from assessing a limited number of structures at a time towards screening entire databases for evaluating the performance of hundreds to millions of solid sorbents. In some of these studies, simple metrics such as selectivity were used to evaluate the performance of the materials. In a more recent study, a vacuum swing adsorption process was developed and materials were ranked based on process indicators such as purity, recovery, parasitic energy, and productivity. These screening studies consider mainly binary mixtures of CO₂ and N₂ in the gas stream. Therefore, the accurate prediction of binary adsorption data plays a key role in evaluating the performance of materials in carbon capture processes. Consequently, the development of reliable and accurate tools to predict mixture adsorption data is required to link materials’ adsorption properties to their process performance.

So far, in computational screening studies, the prediction of mixture adsorption isotherms is usually done by fitting the simulated isotherms for the pure components to one specific mixture adsorption model. Moreover, for process optimization purposes, mixture adsorption data is required at many different operating conditions. It is therefore important that one selects the right adsorption isotherm model such that the data can be described sufficiently accurately. Likewise, if experimental data is used to describe the thermodynamics of an adsorption process, a lot of effort is spent in fitting that data to the right model. This cumbersome fitting process can not be utilized on screening studies of hundreds or thousands of
materials, hence the need for a robust workflow that gives us an optimal prediction of the mixture adsorption isotherms for a wide range of pressures, temperatures, and compositions. In this workflow, we aim to achieve a sufficiently accurate description of the thermodynamics of the adsorption process, with a minimal number of simulation data for the pure components. This is a crucial step because the accuracy of our predictions heavily relies on the quality of the pure component isotherms, especially at low coverage. This low coverage regime is of particular importance for capturing CO$_2$ from dilute sources such as confined spaces or directly from the air, where the CO$_2$ concentration falls significantly below 1%. Several studies have shown that IAST accurately predicts binary CO$_2$ and N$_2$ adsorption isotherms for zeolites and MOFs using computational data. In contrast, Gharagheizi and Sholl concluded from a set of experimental mixture data that for CO$_2$/N$_2$ mixtures, IAST predictions are the worst of all mixtures considered in their study. These contradicting arguments motivated us to carry out a careful study on the reliability of predicting mixture adsorption isotherms at conditions of interest for carbon capture applications.

In this paper, we present a workflow that is designed to work without manual intervention to efficiently predict, by using molecular simulations, the thermodynamic data that is needed to design a carbon capture process. We developed a procedure that does not rely on fitting the adsorption isotherms. From molecular simulations, we can obtain accurate data for both, the pure component isotherms as well as the mixture isotherms. This allowed us to make a detailed comparison of the different methods to predict mixture isotherms. All approaches rely on an accurate description of the pure component isotherms and a model to predict the mixture isotherms. As we are interested in low CO$_2$ concentrations, it is essential that these models correctly predict the low-pressure limit, i.e., give a correct description of the Henry regime. Among the equations that describe this limit correctly, the dual-site Langmuir (DSL) model is often used for the pure components and the extended DSL (EDSL) for the mixtures. An alternative approach, which avoids describing the pure component isotherms with a model, is to numerically integrate the pure component isotherms in the context of
In this work, we compare these two methods. In addition, we show that the way these data are fitted for DSL can significantly impact the ranking of materials, in particular for capture processes with a low concentration of CO$_2$ in the feed stream.

**Simulation details**

In this work, we focus on metal-organic frameworks (MOFs). MOFs are crystalline materials and by changing the linker one can potentially synthesize over a trillion different materials. A lot of research has been focused on finding the optimal MOF for carbon capture.\(^{25-28}\)

Ideally, one would like to have a set of experimental pure components and mixture isotherms to test the approach we developed here. However, the number of MOF structures for which accurate mixture adsorption isotherms have been experimentally measured over a large range of pressures, temperatures, and compositions is very limited, especially for a mixture of CO$_2$ and N$_2$.\(^{29}\) Therefore, we rely on molecular simulations to predict the adsorption isotherms of both, the pure components and their mixture. Here, we use a set of 500 MOF structures that were chosen to represent a diverse set of metals, linkers, ligands, pore sizes, and topologies.\(^{30}\)

In our molecular simulations, we assumed that the crystal structures are rigid. First, each crystal structure was cleaned (i.e., atomic overlaps were removed, solvent molecules were removed, and missing hydrogen atoms have been added) and subjected to Density functional theory (DFT) geometry optimization. The DFT electronic density was also used to compute partial charges via the DDEC protocol.\(^{31}\)

The TraPPE\(^{32}\) force-field was used to describe the CO$_2$-CO$_2$, N$_2$-N$_2$ and CO$_2$-N$_2$ interactions, whereas for the guest-MOF interactions the TraPPE-UFF\(^{33}\) force-field combination was the selected one, as these are common force fields adopted by the community.

We used Monte Carlo simulation in the canonical ensemble of one molecule to compute the heat of adsorption at infinite dilution, $\Delta H_H$ (kJ mol$^{-1}$), and Widom test particle insertions
to compute the Henry coefficient, $K_H$ (mol kg$^{-1}$ bar$^{-1}$). Complete adsorption isotherms are computed using Grand Canonical Monte Carlo simulations (GCMC).

All the calculations were run using the Automated Interactive Infrastructure and Database for Computational Science, AiiDA. Pure component isotherms were obtained using the associated AiiDA workflow named *IsothermAccurateWorkChain*, which is part of the *aiida-lsmo* plugin. Being part of the same plugin, *MulticompGcmcWorkChain* was the workflow used to calculate the binary GCMC data.

In addition to all the isotherms calculations, the heat capacities of the different MOF structures were estimated using our newly developed machine learning method, which simply uses the crystal structure as an input. These heat capacities are used later on as inputs to the process model. More details on the structures, the force fields, and simulations can be found on the Materials Cloud and the SI.

**Pure component isotherms**

If one screens a very large number of structures, it is not feasible to manually decide on the pressures for which the loading is computed from a molecular simulation. One can, of course, use a brute-force method, in which we simulate such a large number of pressure points that we are guaranteed to have a correct sampling for all materials. However, with limited computational resources, it is important to develop a workflow that ensures an efficient sampling for all materials.

**The Computational Workflow**

We carry out the computation of an isotherm in two steps. In the first step, we compute the Henry coefficient of all materials. Only for those materials that have a sufficiently high Henry coefficient for CO$_2$, we will compute the complete isotherm. In the Henry regime, the
The isotherm follows from

\[ q_H = K_H P_H, \quad (1) \]

where \( q_H \) is the loading (mol kg\(^{-1} \)), \( K_H \) is the Henry coefficient and \( P_H \) is the pressure for which we start to observe significant deviations from the Henry regime. As we have computed the Henry coefficient accurately for all materials, we only need to simulate the complete adsorption isotherm from the pressure \( (P_H) \). To estimate \( P_H \), we used the following steps:

- As a first guess of the limit of the Henry regime, we assume to have reached \( P_H \) at a loading of one molecule per unit cell of the crystal. If Henry’s law would hold, this pressure follows from:

\[ \frac{1}{N_A V_{ads} \rho_{ads}} = K_H P_H, \quad (2) \]

where \( V_{ads} \) is the volume of the unit cell in \( (m^3) \), \( \rho_{ads} \) is the density of the unit cell in \( (kg \ m^{-3}) \), and \( N_A \) is Avogadro’s number.

- The true uptake, \( q_H \), is obtained by running a GCMC simulation at \( P_H \). If \( P_H \) belongs to the Henry regime, the error between the true uptake and Henry’s uptake should be smaller than a set precision value, \( \epsilon \), as:

\[ \frac{q_H - P_H K_H}{q_H} < \epsilon. \quad (3) \]

- If the error is higher than \( \epsilon \), we have overestimated the pressure for which the Henry regime holds, and \( P_H \) is decreased by a factor of 0.8. And, we repeat the previous steps until the error converges to a value smaller than \( \epsilon \).

It is important to predict Henry’s coefficients with high accuracy. However, we would like to avoid running the protocol described by equation \( 3 \) indefinitely. Therefore, we allow for a 7.5% error margin compared to the ones obtained by running the Widom Insertion technique.
For pressures above $P_H$, we use GCMC simulations to compute the isotherms. If we approach the saturation loading, these GCMC simulations become very time-consuming as the probability to insert an additional molecule becomes extremely low. In addition, under these conditions, the pressure is much higher than we typically reach in practical applications. Therefore, we stop our GCMC simulations at a pressure $P_{\text{sat}}$, where we have theoretically reached 95% of the estimated maximum loading. We have tested higher loading settings, but, except for increasing the CPU requirements, a higher value did not significantly change the results.

To estimate $P_{\text{sat}}$, we assume that we can describe the adsorption isotherm with a Langmuir isotherm:

$$q = q_{\text{sat}} \frac{bP}{1 + bP}, \quad (4)$$

where $b$ is the adsorption equilibrium constant and can be related to the Henry coefficient. For $P \to 0$, the Langmuir model describes the Henry regime, hence:

$$K_H = q_{\text{sat}} b. \quad (5)$$

At 95% of the maximum loading in the Langmuir approximation gives:

$$0.95q_{\text{sat}} = q_{\text{sat}} \frac{bP_{\text{sat}}}{1 + bP_{\text{sat}}}. \quad (6)$$

Rearranging the equation to solve for $P_{\text{sat}}$ gives:

$$P_{\text{sat}} = \frac{19}{b} = 19 \frac{q_{\text{sat}}}{K_H}, \quad (7)$$

where $K_H$ has been computed previously. To compute $q_{\text{sat}}$, we use the procedure described by Ongari et al. In this procedure, it is assumed that at saturation, the density of adsorbed
molecules, $\rho_{\text{liquid}}^{\text{guest}}$, is equal to the liquid density:

$$q_{\text{sat}} = V_{\text{pore}}^{\text{guest}}/\rho_{\text{liquid}}$$

(8)

where $V_{\text{frame}}^{\text{pore}}$ (m$^3$/kg$^{-1}$) is the probe-occupiable accessible pore volume of the crystal structure and can be computed from the crystal structure by Zeo++.

For the liquid densities of the guest molecules, we used 25.02 and 28.84 (mol m$^{-3}$) for CO$_2$ and N$_2$, respectively. By substituting equation 8 into equation 7 we can compute $P_{\text{sat}}$ using:

$$P_{\text{sat}} = \frac{19V_{\text{pore}}^{\text{guest}}/\rho_{\text{liquid}}}{K_H}.$$  

(9)

GCMC simulations are now performed only on the relevant part of the isotherm, i.e., from $P_H$ to $P_{\text{sat}}$. A further gain in computing efficiency can be obtained if we adjust the number of pressure points to the steepness of the isotherm; the steeper the isotherm the smaller the steps in pressure. To achieve this, the pressure points are selected based on the formula:

$$P_{i+1} = P_i + \min \left( \Delta P_{\text{max}}, \frac{S}{100} \frac{q_i}{\left( \frac{\partial q}{\partial P} \right)_{P_i}} \right).$$  

(10)

The second term, containing the derivative, ensures that the change in uptake does not exceed $S\%$. $S$ can be selected based on how smooth the sampling must be. The higher the value of $S$, the lower the number of pressure points. The maximum pressure step ($\Delta P_{\text{max}}$) is determined so that in the worst-case scenario $N$ points are generated as:

$$\Delta P_{\text{max}} = \frac{P_{\text{sat}} - P_H}{N}.$$  

(11)

The derivative of uptake with respect to pressure is obtained numerically by applying the backward divided difference scheme:

$$\left( \frac{\partial q}{\partial P} \right)_{P_i} = \frac{q_i - q_{i-1}}{P_i - P_{i-1}}.$$  

(12)
The uptake is obtained by running GCMC calculations for each pressure point. The heat of adsorption is also an output of these calculations. Given that the uptake is an increasing function of pressure, the derivative must always be positive. In the case that this is not true, the last run at $P_i$ is always repeated.

Figure 1 shows the pure component isotherms of CO$_2$ and N$_2$ computed at 25°C for an exemplar group of 10 structures. In the case of CO$_2$, the workflow successfully predicted saturation, where the reduced uptake is shown to converge to 1, i.e., plateuing when reduced pressure approaches 1. In the case of N$_2$, higher pressures are required to reach saturation. However, $P_{sat}$ values are beyond interest and it has no additional value to continue further. Finally, we use $N = 25$, hence 25 points per isotherm to ensure a smooth sampling of the pressure points. The workflow was later tested on the original set of 500 structures and was found to be reliable irrespective of the structure or the guest molecule, i.e., CO$_2$ and N$_2$, under investigation.

Figure 1: Pure component isotherms at 25°C for CO$_2$ (left) and N$_2$ (right). The plots show the reduced uptake as a function of reduced pressure for the exemplar group of 10 structures.
Clausius-Clapeyron Relation

In pure component adsorption, two phases coexist at equilibrium in the system: the (ideal) gas phase and the adsorbed phase. In the same way, the Clausius-Clapeyron relation is used to relate temperature and pressure along phase boundaries in a phase diagram, it can also relate temperature and pressure for pure component adsorption systems. For adsorption, the Clausius-Clapeyron relation states that for two different isotherms, \( T_{\text{ref}} \) and \( T_{\text{new}} \), the same loading is obtained at two different pressures \( P_{\text{ref}}^i \) and \( P_{\text{new}}^i \) respectively:

\[
P_{\text{new}}^i = P_{\text{ref}}^i \exp \left( \frac{-\Delta H_{\text{ads}}^{\text{ref}}}{R} \left( \frac{1}{T_{\text{new}}} - \frac{1}{T_{\text{ref}}} \right) \right),
\]

(13)

where \( \Delta H_{\text{ads}}^{\text{ref}} \) is the heat of adsorption at \( T_{\text{ref}} \) and \( P_{\text{ref}}^i \). Hence, the Clausius-Clapeyron relation can be used to predict pure component isotherms at different temperatures, given a reference isotherm. At the reference temperature \( T_{\text{ref}} \), our GCMC simulations at a pressure point \( P_{\text{ref}}^i \), result in the corresponding uptake, \( q_{\text{ref}}^i \), and heat of adsorption, \( \Delta H_{\text{ads}}^{\text{ref}} \). The Clausius-Clapeyron relation allows us to compute at which pressure, \( P_{\text{new}}^i \), the same amount is adsorbed (\( q_{\text{ref}}^i = q_{\text{new}}^i \)) but at a different isotherm, \( T_{\text{new}} \). Refer to fig [S1] for a visual description.

In equation [13] we have assumed that the heat of adsorption is a constant with respect to temperature, which is a valid assumption for a small temperature difference with respect to the reference temperature, i.e. 25°C. In the SI, we choose a subset of 50 structures, in which we made sure to have a variety of metals, linkers, and pore sizes. We use Clausius-Clapeyron to extrapolate to four different temperatures: 50, 75, 100, and 125°C. Then we use IAST to predict binary adsorption data of \( \text{CO}_2 \) and \( \text{N}_2 \) at those temperatures, for a wide range of \( \text{CO}_2 \) compositions (0.04%-95%). Finally, the predictions are compared to the binary data computed using the pure component isotherms generated by GCMC calculations at the respective temperatures. We were able to show that if we compute the reference \( \text{CO}_2 \) and \( \text{N}_2 \) isotherms at 25°C, we can obtain, for the majority of the predictions, binary \( \text{CO}_2 \) and \( \text{N}_2 \) data...
within an accuracy of 1% to 3%. The higher the temperature difference with respect to the reference temperature, i.e. 25°C, the higher the errors, which corroborates the assumption that a constant heat of adsorption is only valid for small temperature differences. In the extreme case of 125°C, very few predictions were calculated with errors up to 25%. One way to overcome this issue is to use another reference isotherm at a relatively higher temperature, 75°C for example, in order to minimize the temperature difference when extrapolating at high temperatures.

**Binary Adsorption Isotherms**

The number of published experimental binary CO₂ and N₂ adsorption isotherms in MOFs are very limited. For a pure component isotherm, one can measure the weight increase as a function of pressure, while for a mixture, one also has to measure the relative concentration in the pores. Hence, for most practical applications we have to rely on the predictions of binary adsorption isotherms using the pure component isotherms as input.

Unlike experiments, in a GCMC simulation, the computation of a binary adsorption isotherm is similar to a pure component isotherm. But, for most practical applications one has to determine the isotherms for (many) different compositions, making this a computationally expensive step.

In the literature, different routes are used to predict binary adsorption isotherms. In this section, we give a short summary of the different methods. For a more extensive review we refer to the literature.

**Ideal Adsorbed Solution Theory**

Ideal Adsorbed Solution Theory (IAST) is a thermodynamic theory that is applied in this study for predicting mixture adsorption data from calculated pure component isotherm data at a given temperature. The method relies heavily on the thermodynamics of physical
adsorption, and its accuracy depends on four main assumptions:

- Adsorbates have equal access to the entire surface area of the adsorbent.
- During the adsorption process, changes in the thermodynamic properties of the framework are negligible compared to the property changes of the adsorbable species.
- The adsorbed phase is described by the Gibbs energy of adsorption. One simply substitutes the spreading pressure ($\pi$) for pressure ($P$) and area ($A$) for volume ($V$):

$$dG = -SdT + Ad\pi + \sum \mu_i dn_i. \quad (14)$$

- The vapor phase is considered an ideal gas and the adsorbed phase an ideal solution.

With these assumptions, the adsorbed gasses can be described as an ideal gas mixture in equilibrium with an ideal adsorbed phase on the adsorbent surface. For a binary mixture of, say, CO$_2$ and N$_2$, and for a given temperature, gas composition, and the total gas pressure, the number of moles adsorbed for a binary mixture is obtained by solving simultaneously the set of equations governing IAST. One of these IAST equations follows from the condition that at equilibrium, the spreading pressures of CO$_2$ and N$_2$ are equal:

$$\pi_{CO_2}(P^0_{CO_2}) = \pi_{N_2}(P^0_{N_2}) \quad (15a)$$

$$\frac{\int_0^{P^0_{CO_2}} q^0_{CO_2}(P) \, dP}{P} = \frac{\int_0^{P^0_{N_2}} q^0_{N_2}(P) \, dP}{P} \quad (15b)$$

$$P^0_i = \frac{P y_i}{x_i} \quad (15c)$$

Here $q^0_i$ is the pure component isotherm, $P$ is the total pressure, $y_i$ and $x_i$ are the gas phase and adsorbed phase compositions respectively, and $i$ can be either CO$_2$ or N$_2$. Finding a solution for (Eq. 15b) involves solving integrals. Mathematically, integrals can be solved either analytically, if the pure component isotherm has a functional form ($q^0_i = f(P)$),
or numerically if it does not. Simon et al. developed a python package, Ideal adsorbed solution theory Python package (pyIAST), that supports both options. Pure component isotherms can either be fitted to thermodynamically consistent models such as Langmuir, Dual-site Langmuir, Quadratic, BET, Henry’s law, and approximated Temkin, or be linearly interpolated in order to calculate the integral using numerical quadrature. In this work, we use the numerical option of pyIAST. The numerical option has the practical advantage that the outcome does not depend on the functional form that is used to describe the experimental or computational adsorption isotherm. This agrees with the work of Chen and Sholl, where they introduce a new Monte Carlo technique and combine it with IAST to avoid curve fitting.

Extended Dual Site Langmuir Model

The DSL isotherm model was introduced in an attempt to describe non-ideal adsorbate-adsorbent interactions. In this model, it is assumed that the deviations from ideality are due to surface heterogeneity. The model consists of two sites, 1 and 2, that are independent of each other, where each site follows the Langmuir equation. Hence, the dual site adsorption of a pure component is simply the sum of adsorption on each site:

\[ q = q_{sat1} \frac{b_1 P}{1 + b_1 P} + q_{sat2} \frac{b_2 P}{1 + b_2 P}. \]

(16)

It is important to note that this model meets the thermodynamic consistency requirements for pure gas adsorption, i.e., at infinite dilution, the limiting slope of the isotherm is positive and finite, the spreading pressure is a continuous and well-behaved function of pressure, and the isosteric heat of adsorption at constant uptake is finite.

The DSL model can be extended to predict a binary mixture isotherm. If we consider a mixture of gasses A and B, the corresponding extended DSL model for a binary mixture is
described as\cite{23}

\begin{align}
q_A &= q_{\text{sat}1, A} \frac{b_1, A P y_A}{1 + b_1, A P y_A + b_1, B P y_B} + q_{\text{sat}2, A} \frac{b_2, A P y_A}{1 + b_2, A P y_A + b_2, B P y_B}, \\
q_B &= q_{\text{sat}1, B} \frac{b_1, B P y_B}{1 + b_1, A P y_A + b_1, B P y_B} + q_{\text{sat}2, B} \frac{b_2, B P y_B}{1 + b_2, A P y_A + b_2, B P y_B}.
\end{align}

(17a) \hspace{1cm} (17b)

It is important to note that for the extended model to be thermodynamically consistent, the saturation capacities for both components must be equal, therefore:\cite{23}

\begin{align}
q_{\text{sat}1, A} &= q_{\text{sat}1, B}, \\
q_{\text{sat}2, A} &= q_{\text{sat}2, B}.
\end{align}

(18a) \hspace{1cm} (18b)

To generate the binary adsorption data, the parameters of the extended Langmuir model are obtained from the fitting of the pure component isotherms to the DSL model. In some cases, the number of data points within the pure component isotherm is insufficient for fitting purposes. To generate more data points, we applied linear interpolation on the original pure component isotherms.

The parameters of the DSL model are sensitive to the fitting procedure used. Consequently, process modeling results will also be affected\cite{17} We followed the fitting procedure discussed by Farmahini et al.\cite{47} to ensure thermodynamic consistency, which involved minimizing the least-square error between the simulated isotherms and the ones predicted by the model.

**Results**

For the pure component isotherms, we observed that the fitting procedure to obtain the DSL parameters is sensitive to initial guesses, i.e., different initial guesses lead to different model parameters. For example, figure S4 shows the results of exemplar fits. In the region
where we have abundant data, the different fits give an equally good description. This is not surprising as many combinations of parameters give comparable results. However, for applications where CO$_2$ concentration is very dilute, it is important that the model accurately covers the Henry regime. Figure S4 also shows that the different initial guesses give very different predictions for the low-pressure regime. In some cases, it was impossible to obtain a meaningful fit that covers both low and high-pressure regions simultaneously. Therefore, the binary predictions at low or high pressures, depending on where we put the weight on the fit, will be compromised (see figure S5). In addition, if we fit the data we are not guaranteed to obtain values of the DSL parameters that are physically reasonable. For example, in equation (16) $q_{\text{sat}1}$ and $q_{\text{sat}2}$ are directly related to the saturation uptakes, but the values obtained from the fits do not match the ones from the pure component isotherms. In figure S6 we show how different these values are. Higher differences are observed in structures for which the saturation takes place way beyond 10 bars. Therefore, the fitting is sensitive to the range of data being used.

In a few cases, the fitting procedure completely fails or the obtained parameters are physically inexplicable. The predicted saturation uptake is in the order of a 1,000 mol kg$^{-1}$, or the heat of adsorption is a positive value. In those cases, the pure component isotherms were accurately described by the DSL model mathematically, however, thermodynamically, the parameters are not valid, and the fits cannot be trusted (see figure S7). Generally, not every isotherm can be fitted to a DSL model, and assuming that in a set of pure component isotherms for different materials, all isotherms follow the DSL model, or any other model, can be erroneous.

In all these cases, we also compared the predictions of the mixture isotherms with simulation results. The results are shown in figure S8. The direct integration of the pure component isotherms in the IAST framework avoids all these fitting problems and gives reasonable results for the mixture isotherms.

To quantify the differences in accuracy of the predictions between IAST and extended
DSL, we compared the predicted and simulated (GCMC) CO$_2$ and N$_2$ mixture isotherms for a range of temperatures (25 - 125°C) and CO$_2$ compositions (0.0004 - 0.95), and a fixed pressure (1.01325 bar) for the subset of 50 structures. All data points were put together, and we used box plots to visualize the error distribution as shown in figure 2.

Figure 2: Relative error distribution for all conditions and 50 structures. The relative errors were obtained using $RE_i = 100 \left| \frac{q_{i,IAST/DSDL} - q_{i,GCMC}}{q_{i,GCMC}} \right|$. In these box plots, the middle line in the box divides the data set in two: 50% of the data has an error smaller than the one corresponding to the line and 50% an error larger. The region between the lowest whisker and the bottom of the box contains the 25% of the data with the lowest error, while the top of the box and the upper whisker contains the 25% of the data with the highest error. These box plots are shown without outliers, i.e., data points that are located outside the whiskers of the box plot.

When comparing the relative error, IAST predictions were closer to the binary GCMC calculations than the predictions given by the extended DSL model. This was true for all the different conditions, structures, and guest molecules. One can see that the median relative error for both molecules, using IAST, is less than 5% compared to 10% when using
extended DSL. In each case, we had 1250 data points and the outliers counted for almost 10%. These outliers represent the data lying outside the whiskers of the box plot. However, some of them are due to inaccurate binary GCMC calculations, which are usually at very low pressures. From these results, one can conclude that IAST provides good predictions of binary adsorption data for the set of diverse structures over the temperature, and CO$_2$ partial pressure range of interest. These results are of no surprise. It was previously established, in multiple studies,\textsuperscript{14,18,19} that IAST accurately predicts binary CO$_2$ and N$_2$ adsorption isotherms for zeolites and MOFs using computational data.

**Importance of Isotherm Sampling**

Gharagheizi and Sholl\textsuperscript{21} concluded in their study of the reliability of predictions of mixture isotherms that for a CO$_2$/N$_2$ mixture, IAST gave the worst results. This conclusion appears to be in sharp contrast with our conclusion that IAST performs the best. In this section, we show that the work of Gharagheizi and Sholl points at an important limitation of the IAST theory, which becomes apparent if one uses the pyIAST\textsuperscript{24} package.

The pyIAST package was developed with the idea that mixture isotherms are computed directly from the experimental (or simulated) pure component isotherms, without the need for any fitting. In our workflow to compute the pure component isotherms, we have ensured that we obtain adsorption isotherms up to 90% of the saturation loading, where we estimate the saturation loading from the pore volume. However, experimentally, in particular for N$_2$, one often has isotherms that have been measured up to a maximum pressure that is far from saturation. In the case of such an incomplete isotherm, one has to be careful with pyIAST. The advantage of the pyIAST package is that we obtain the optimal IAST prediction given the experimental (or computational) data that are provided, and we do not have to rely on a fitting step. However, this also implies that pyIAST is unable to extrapolate to pressures higher than are provided. pyIAST extrapolates beyond the last pressure point by assuming that saturation was reached, and the loading is a constant, equal to the largest loading
available in the pure component data. The pyIAST gives a warning that it had to use extrapolated data, but it can not give an indication of the error this might give. This is best illustrated in the left-hand side of Figure 3.

Given that IAST calculations are performed using numerical quadrature to compute the spreading pressure, it is important to understand how “extrapolating” beyond the last pressure point using pyIAST affects the mixture predictions. On the right-hand side of Figure 3 we consider a complete pure component isotherm (red line), which is then truncated at $P_{cut}$. pyIAST is used to extrapolate the resulting incomplete isotherm (blue line) as described previously. The shaded area corresponds to the integration difference between the complete and incomplete isotherms and can be described by:

$$\int_{P_{cut}}^{P_i^0} \frac{q_i^0(P)}{P} dP - \int_{P_{cut}}^{P_i^0} \frac{q_{i,\text{sat}}}{P} dP = \int_{P_{cut}}^{P_i^0} \frac{q_i^0(P)}{P} dP - q_{i,\text{sat}} \ln \frac{P_i^0}{P_{cut}} \quad \text{(19)}$$

Where $P_i^0$ is previously defined in equation 15c and $i$ can be either CO$_2$ or N$_2$. By definition, equation 19 is only valid for $P_{cut} < P_i^0$.

Figure 3: Extrapolating beyond the pressure limit for an incomplete isotherm using pyIAST (left). The difference between a complete and an extrapolated incomplete isotherm (right).
To obtain some insights on the importance of this truncation, we choose pure component isotherms from one of the exemplar structures shown in Figure 1. We mimic the case in which we can only have data up to maximum pressure, which is a fraction of the saturation pressure, by truncating the corresponding CO₂ and N₂ pure component isotherms at different pressures. Those isotherms are then used along with pyIAST to predict binary data at 25°C, 1.01325 bar and different gas compositions (10% and 90% CO₂). The binary data predicted using the extrapolated incomplete isotherms is compared to the binary data predicted using the complete pure component isotherms, by computing the relative error for CO₂ and N₂.

The saturation uptake of the structure in question is 4.7 mmol g⁻¹, and is reached at 2 bar and 246 bar for CO₂ and N₂, respectively. Using the complete pure component isotherms, we predict $P^0_{CO₂}$ and $P^0_{N₂}$ to be equal to 0.1 bar and 14 bar respectively for a mixture of 10% CO₂. For a mixture of 90% CO₂, $P^0_{CO₂}$ and $P^0_{N₂}$ are equal to 0.9 bar and 126 bar respectively. We observe that increasing the composition of CO₂ will increase the $P^0_i$ values by a similar factor, in this case by a factor of approximately 9.

Figures 4 and 5 show the effect of truncating the CO₂ isotherm on CO₂ IAST predictions keeping the N₂ pure component isotherm intact (left), as well as the effect of truncating the N₂ isotherm on N₂ IAST predictions keeping the CO₂ pure component isotherm intact (right) for a mixture of 10% and 90% CO₂, respectively. It is evident from both figures that, if the pure component isotherm of component i is truncated at a pressure higher than $P^0_i$, the IAST predictions are exact and the relative error is always null. This follows from the definition of equation 19. Nevertheless, if the truncated pressure lies to the left of $P^0_i$, the pure component isotherm looses "important information" and pyIAST’s extrapolation leads to incorrect mixture predictions. Also, the further we truncate to the left, the higher the relative error. In fact, there is an exponential correlation between the relative error and the truncated pressure.
Figure 4: The plots show the relative error in IAST predictions as a function of the truncated pressure for CO$_2$ (left) and N$_2$ (right) for a mixture of 10% CO$_2$. The dashed lines correspond to $P^0_i$ values at the given mixture composition.

Figure 5: The plots show the relative error in IAST predictions as a function of the truncated pressure for CO$_2$ (left) and N$_2$ (right) for a mixture of 90% CO$_2$. The dashed lines correspond to $P^0_i$ values at the given mixture composition.

The value of $P^0$ dictates how accurate the sampling of a pure component isotherm should
be. However, this value cannot be predicted a priori. Our computational workflow offers a suitable solution and can be safely used in conjunction with pyIAST, as it generates CO$_2$ and N$_2$ isotherms up to the saturation point. Interestingly, N$_2$ saturation is reached at relatively higher pressures compared to CO$_2$. From an experimental perspective, complete N$_2$ isotherms will be difficult to obtain due to the equipment limitations.

We can now understand why Gharagheizi and Sholl$^{21}$ obtained the worst results with IAST, as they implicitly assumed by using pyIAST that the experimental N$_2$ were obtained at sufficiently high pressures. However, most experimental N$_2$ do not reach this pressure. If this is the case assuming a constant loading beyond the maximum pressure that is measured, leads to large errors.

IAST predicts the mixture isotherms surprisingly accurate, but it can only do this at all compositions if it “knows” the pure component isotherms up to sufficiently high pressures. Fortunately, it is possible to extrapolate to these higher pressures if we know the pore volume of the material (e.g., from BET calculations), then use equation 8 to approximate the saturation loading. With this approximation of the saturation loading and the low pressure experimental data, one can make a sensible extrapolation of the experimental data. If these data are then added to the experimental data, (py)IAST will most likely give a much better prediction. Unfortunately, N$_2$ isotherms are often overlooked and people tend to put more emphasis on CO$_2$ isotherms.

**Process Modeling**

In the previous section, we have shown that IAST gives a better prediction of the mixture adsorption isotherms compared to the extended DSL. In an industrial process simulation, the use of IAST requires much longer CPU times because of the integration, and because of this, preference is given to extended DSL or related equations which are numerically less demanding. These CPU times get even longer when we consider optimizing the process,
given the multiple iterations involved. However, very little is known about how these differences in the thermodynamic description would impact the process design. For example, would a ranking of the performance of materials significantly change, or is the impact of these differences relatively minor? To answer this question, we consider an elementary TSA process reproduced from the work of Ajenifuja et al.\cite{Ajenifuja2020} to capture CO\textsubscript{2} from three different sources: (1) a coal-fired power plant (CF-PP), (2) a natural gas power plant (NG-PP), and (3) confined spaces (CS). This process is shown schematically in figure 6.

![Figure 6: A standard 3-step cycle TSA process: (i) adsorption, (ii) open heating, and (iii) open cooling](image)

In evaluating the process performance, we considered two distinct fitting schemes, DSL and DSL\textsubscript{Henry} respectively. In the first, all data points were given the same weight. Whereas, in the second, we put more weight on the low-pressure region, such that we force the DSL equation to describe the Henry coefficient correctly. In figure 7 we show the materials ranking for a coal-fired power plant (CF-PP), natural gas-fired power plant (NG-PP), and confined spaces (CS). In these figures, we compare the purity (%) of CO\textsubscript{2} rich product stream and we
use the IAST results as the reference. For a fair comparison between IAST and extended DSL, 100 out of the initial 500 structures, for which the fitting procedure failed or the obtained parameters were physically inexplicable, were labeled as "problematic" and were discarded.

![Figure 7: Materials ranking in terms of purity of the output stream for a coal-fired power plant (CF-PP), natural gas-fired power plant (NG-PP), and confined spaces (CS). In these figures, each circle corresponds to a material, where the middle gives the ranking as predicted by IAST, the left and right from DSL, in which the right side more emphasis is put on fitting the Henry coefficient correctly. Points that correspond to the same material are connected by lines, where a gray line is used if the change between IAST and DSL or DSLHenry is less than 10%, a blue line if the DSL gives 10% or higher purity, and red if it is 10% or lower.](image)

As can be expected, the range of purity that can be obtained in the process depends on the concentration of CO₂ in the inlet gas and, in this case and for the evaluated TSA process, purity is lowest for the confined spaces. This is best illustrated in figure 7. If we focus on the ranking of the materials in figure 7 we see for the coal case that the top-performing materials all give a similar purity, independently of the method that is used. However, if we go to natural gas or confined spaces, the order of the materials does depend on the method that is used, which corroborates the conclusion of the previous section that at low concentrations the differences between DSL and IAST can be significant. Interestingly, if we
give more weight to the low-pressure regime we tend to underestimate the purity (more red lines) and without this weight, we overestimate (more blue). However, this trend is not true for all structures. Similar conclusions hold for other key performance indicators, such as the working capacity and specific heat requirements (see section 3.1 in the SI).

One important advantage of IAST is that we do not need to fit the data, while DSL does depend on how the data are fitted. In figure 8 we illustrate the effect of the fitting procedure on the ranking. These figures show that DSL tends to give a higher purity than DSLHenry, but this does not hold for all structures.

![Figure 8: Materials ranking in terms of purity of the output stream for a coal-fired power plant (CF-PP), natural gas-fired power plant (NG-PP), and confined spaces (CS). In these figures, we compare DSL with DSLHenry, see also the caption to figure 7.](image)

Another important aspect that is worth highlighting is that, in our simulations, we can obtain accurate data at very low pressure. Experimentally, however, the low-pressure regime is limited by the specifications of the available equipment in the laboratory. We use our data to estimate the impact of these experimental limitations in the low-pressure regime on the performance. For this, we consider two different scenarios. In the first, we assume that accurate data can be obtained above 5 mbar, and in a second scenario, we assume more
routine experiments for which the minimum threshold was set at 25 mbar. In each scenario, the structures for which the CO$_2$ and N$_2$ reference isotherms did not have data points below the corresponding threshold, the fitting procedure failed or the obtained parameters were physically inexplicable, were discarded. We were left with 95 and 250 structures in each scenario, respectively. For those materials, it is important to realize the consequence of removing data points in the Henry regime. Figures 9 show that the impact on the purity is higher when isotherm data only above 25 mbar was used. Some materials adsorb CO$_2$ very strongly and from those materials 25 mbar is already outside the Henry regime. We can also observe that the lower the CO$_2$ concentration, the more the process performance is affected.
Figure 9: Materials ranking in terms of purity of the output stream for a coal-fired power plant (CF-PP), natural gas-fired power plant (NG-PP), and confined spaces (CS) using isotherm data at pressures above 5 mbar (a) and 25 mbar (b). The middle gives the ranking as predicted by IAST, the left and right from DSL, in which the right side more emphasis is put on fitting the Henry coefficient correctly. See also the caption to figure 7.

We can quantify how the ranking is changed by computing the Spearman coefficient, which is a measure of how much the ranking of the material is impacted. Table I shows the
Spearman coefficients for the different cases. A Spearman coefficient of 1 implies that the ranking has not changed, while 0 indicates that the order has been completely lost.

For a coal-fired power plant, we observe that the ranking of the materials is little affected by the method we use (Spearman coefficient larger than 0.85). Only if we lack isotherm data below 5 and 25 mbar and impose the fitting of the Henry coefficient, errors in our ranking become apparent (Spearman coefficient of 0.79 and 0.74, respectively). This is for cases for which we impose an incorrect Henry coefficient as the data for 5 and 25 mbar are already outside the Henry regime. For the natural gas case and confined spaces, the concentration of CO$_2$ is so low that purity values are very sensitive to the method we use. However, the ranking of the material does not show significant deviations (Spearman coefficient larger than 0.8). Only in the case of confined spaces, if we lack isotherm data below 5 and 25 mbar, errors in our ranking become apparent (Spearman coefficient of 0.78 and 0.77 respectively).

Table 1: The Spearman coefficient comparing the ranking of materials between DSL fitting and IAST for a coal-fired power plant (CF-PP), natural gas-fired power plant (NG-PP), and confined spaces (CS). The subscripts “5” and “25” indicate that the isotherm data below 5 and 25 mbar, respectively, have been removed from the data set to mimic the experimental limitations. The subscript “Henry” indicates that the fitting is done to ensure that the fits capture the Henry coefficient.

<table>
<thead>
<tr>
<th></th>
<th>DSL</th>
<th>DSL$_{Henry}$</th>
<th>DSL$_5$</th>
<th>DSL$_{5Henry}$</th>
<th>DSL$_{25}$</th>
<th>DSL$_{25Henry}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF-PP</td>
<td>0.91</td>
<td>0.85</td>
<td>0.84</td>
<td>0.79</td>
<td>0.89</td>
<td>0.74</td>
</tr>
<tr>
<td>NG-PP</td>
<td>0.86</td>
<td>0.81</td>
<td>0.86</td>
<td>0.85</td>
<td>0.81</td>
<td>0.86</td>
</tr>
<tr>
<td>CS</td>
<td>0.86</td>
<td>0.81</td>
<td>0.78</td>
<td>0.83</td>
<td>0.77</td>
<td>0.86</td>
</tr>
</tbody>
</table>

The differences between DSL and IAST can cause some of the top-performing materials, as predicted by IAST, to not appear at the top for DSL. Table 2 summarizes the number of structures that appear in the top 40 performing materials as predicted by IAST for the different case studies using the DSL fitting. It is evident that one can miss from 10% in the best-case scenario to 33% of the top-performing structures when using extended DSL compared to IAST.
Table 2: The number of structures that appear in the top 40 performing materials as predicted by IAST for a coal-fired power plant (CF-PP), natural gas-fired power plant (NG-PP), and confined spaces (CS) using DSL fitting. See caption to table.

<table>
<thead>
<tr>
<th></th>
<th>DSL</th>
<th>DSL\textsubscript{Henry}</th>
<th>DSL\textsubscript{5}</th>
<th>DSL\textsubscript{5Henry}</th>
<th>DSL\textsubscript{25}</th>
<th>DSL\textsubscript{25Henry}</th>
</tr>
</thead>
<tbody>
<tr>
<td>CF-PP</td>
<td>35</td>
<td>34</td>
<td>35</td>
<td>31</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>NG-PP</td>
<td>29</td>
<td>27</td>
<td>34</td>
<td>35</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>CS</td>
<td>27</td>
<td>28</td>
<td>32</td>
<td>36</td>
<td>33</td>
<td>30</td>
</tr>
</tbody>
</table>

An interesting question is whether IAST has similar difficulties if we remove the data at very low pressure. In figure 10, we make a similar comparison for IAST with computational data removed below 5 and below 25 mbar. In this comparison, one has to realize that we now have IAST predictions on both sides. Hence we are only comparing the effect of omitting the low-pressure data. Whereas, in such a comparison for DSL, we have a convoluted effect as we see the differences between IAST and DSL and, on top of this, the effect of omitting low-pressure data. Yet, the effect is surprisingly small, and even for the confined spaces case, one can obtain a decent ranking. If one uses DSL, one needs to fit the coefficients. Depending on the fitting procedure and set of data, one may get different sets of parameters that optimally fit the data but may give very different results when extrapolating to low pressures. This is also illustrated in figure 9. For coal, we see that the ranking of DSL does not depend on the fitting, while for natural gas and confined spaces, we get a completely different ranking. The IAST procedure does not suffer from this and therefore yields a more reliable extrapolation to low pressures.
Concluding remarks

In this study, we developed a robust workflow that optimally computes pure component isotherms at a given temperature and for a given guest molecule. The workflow was developed and tested for a set of 50 diverse MOF structures and later on, applied to 500 other structures. We showed that IAST can accurately predict binary uptake for a mixture of CO$_2$ and N$_2$, provided that the pure component isotherms are known up to sufficiently low and high pressures. We highlighted the differences between the use of IAST and Extended DSL and were able to prove that IAST as a numerical tool is more robust and accurate than using analytical methods, in this case, DSL. Moreover, our method does not rely on fitting isotherm data to a particular model. This is important for screening a large number of materials where it is nearly impossible to manually inspect the quality of the fit to the data.

We observed that the performance ranking of materials, given a process design, is not
only sensitive to the quality of the pure component isotherms but also to the selected thermodynamic description of the binary uptake data. The ranking of materials for a given key process performance indicator depends on whether we use IAST or DSL.

From a process simulation point of view, predicting mixture isotherms using IAST requires more CPU time than using a method that provides the mixture isotherm in an analytical form (e.g., DSL). This can be problematic, i.e., very computationally expensive, if one needs to optimize the process. Possibilities to address the CPU issue include fitting the IAST mixture predictions to an analytical expression or pre-computing the mixture isotherm on a grid and using interpolation techniques.

Acknowledgement

This work is part of the PrISMa Project (299659), funded through the ACT Programme (Accelerating CCS Technologies, Horizon 2020 Project 294766). Financial contributions from the Department for Business, Energy & Industrial Strategy (BEIS), together with extra funding from the NERC and EPSRC Research Councils, United Kingdom, the Research Council of Norway (RCN), the Swiss Federal Office of Energy (SFOE), and the U.S. Department of Energy are gratefully acknowledged. Additional financial support from TOTAL and Equinor is also gratefully acknowledged. C.C. and S.G. are also supported by the UKRI ISCF Industrial Challenge within the UK Industrial Decarbonisation Research and Innovation Centre (IDRIC) award number EP/V027050/1. S.M.M. acknowledges support from the Swiss National Science Foundation (SNSF) under Grant P2ELP2_195155.

Supporting Information Available

The supporting information gives details on the accuracy of the Clausius-Clapeyron relations and the process modelling. All the details of the simulations (input files, structures, pure component isotherms, and mixture simulations) can be found on the Materials Cloud.
References


