Can we systematically screen millions of chemical structures for cost-effective carbon capture?

¿Podemos cribar sistemáticamente millones de estructuras químicas para una captura de carbono económica?

Charithea Charalambous[1], Fergus Mcilwaine[1], Elias Moubarak[2], Berend Smit[2], Susana Garcia[1]*

[1] Research Centre for Carbon Solutions (RCCS), School of Engineering and Physical Sciences, Heriot-Watt University, EH14 4AS, Edinburgh, UK.

[2] Laboratory of Molecular Simulation (LSMO), Institut des Sciences et Ingénierie Chimiques, Ecole Polytechnique Fédérale de Lausanne (EPFL), Rue de l'Industrie 17, CH-1951 Sion, Switzerland.

** s.garcia@hw.ac.uk*

Abstract

Finding the optimal solid adsorbent to capture CO_2 for a given source of CO_2 and sink (destination) of CO_2 is an interesting scientific and technological question. There are millions of materials to choose from and we lack the capacity to synthesize and test all of them. In this work, we show how one can computationally screen thousands of materials and identify the best performing ones by using process-driven key performance indicators that are representative of a capture process that is specifically design for a particular CO_2 source and sink. As an illustration of the methodology, we use as an example the capture of CO_2 from a waste-to-energy power plant.

Introduction

There is an urgent global effort to harness the momentum of the Paris Agreement and to limit global warming to well below 2 ºC compared to pre-industrial levels. To achieve this, many countries have made commitments to reach net-zero by mid-century. The sense of urgency is apparent: financial and political climate-related investment has dramatically increased, emission mitigation technology is being deployed at an accelerated rate, and researchers focus on finding economically viable solutions to meet the Paris climate targets.

The vast majority of greenhouse gas emissions (CO_2 emissions) can be minimised by either reusing the emitted CO_2, or preventing the CO_2 from being released into the atmosphere through carbon capture and storage. To do so, cost-effective Carbon Capture Utilization and Storage (CCUS) processes are required to be rapidly deployed across a range of different sources of CO_2. The existence of many different CO_2 sources, CO_2 sinks, and economic conditions across the globe requires CCUS technologies to be evaluated on a case by case basis.

Sorbent-based carbon capture is one of the available technologies suitable for the efficient removal of gas impurities aiming to achieve extremely high purities [1]. The two main reasons for this are: (a) the availability of a large spectrum of microporous adsorbents with varying pore structures and surface properties, and (b) the possibility of designing many different process schemes by tailoring generic adsorption separation methods. Materials like metal-organic frameworks (MOFs) are good candidates, as their chemical composition and pore shape can be optimally tuned for particular applications. The challenge here is to effectively screen the large number of available MOFs for their suitability in a carbon capture process.

Usually, a new material is developed with a particular application in mind. There is no indication whether this material could be a good candidate for other applications, unless one could evaluate its performance for every application. With the tools available today, this task is unrealistic. The development of open-access material databases (e.g., CSD, AtomWork, ChemSpider) provides us with the opportunity of efficient data storage, management, query, presentation, and manipulation. Given the wide spread of material development and data storage, there is a recent trend of linking materials to chemical processes for designing cost-effective optimal separation processes. By accessing these large pools of chemical structures and computing their properties, one can evaluate their performance in a particular application. However, these calculations can be very computationally expensive to perform.

Machine learning has been successfully applied to this domain in multiple ways. In certain situations, expensive molecular simulations can be replaced by multipurpose multilayer perceptron (MLP) [2] to predict isotherms. From a process perspective, surrogate models can be developed that mimic detailed carbon capture processes models, requiring only a fraction of the time to compute results [3]. In contrast to screening studies, inverse design can leverage machine learning to generate MOFs with specific properties [4].

Surrogate models require training in order to predict the outcomes of a carbon capture process. A simple case would be to train a surrogate model on a material database to predict key performance indicators (KPIs) for each material. However, selecting the set of materials to train the model is non-trivial. Since only few materials are top performers in many databases, a model trained on such a database is predominantly trained on the poorly performing materials, leading to poor predictions on high performing materials. A more suited approach is to train a model based on a set of materials that give good performance

predictions. This can be done using active learning, where a more detailed process model is queried for interesting materials, the training data is updated, and the surrogate model retrained [5].

The ultimate goal of most screening studies is to determine the top performing materials. In our case, we determine the performance of a material by its KPIs and so it is difficult to identify a top performer without stating which KPI we find more important than others. Instead, we find the Pareto optimal materials, i.e. those set of materials for which no KPI can be improved without loss in another KPI.

Our previous work involved the development of a molecular simulation tool which allowed us to screen thousands of materials for adsorption using their physicochemical and adsorptive properties [6]. Most recently, we extended this work by coupling the molecular simulation tool with process modelling to rank materials for gas separations and for a given set of process metrics, including the overall energy efficiency and process productivity [7]. A 4-step temperature-vacuum swing adsorption (TVSA) process was developed, and its performance was evaluated using an equilibrium-based shortcut model similar, but extended, to that described by Joss et al. [8]. Our TVSA process model is then coupled with molecular simulations in a high-throughput screening platform. The screening platform allows us to evaluate both the physicochemical and adsorptive properties of thousands of microporous structures, including around 300,000 MOFs [6], and their performance for a particular carbon capture application. A simple representation of the screening platform, excluding some of its current features (discussed by Riboldi et al. [7]) for the purpose of this article, is provided in Figure 1.

We compare this approach to one using active learning. We are able to find the same top performers in only a fraction of the time compared to a brute-force technique (i.e., coupling molecular simulations with process modelling to calculate the KPIs of each material). Our strategy uses an adapted version of the ε-PAL method [9], available as a python package named PyePAL [5].

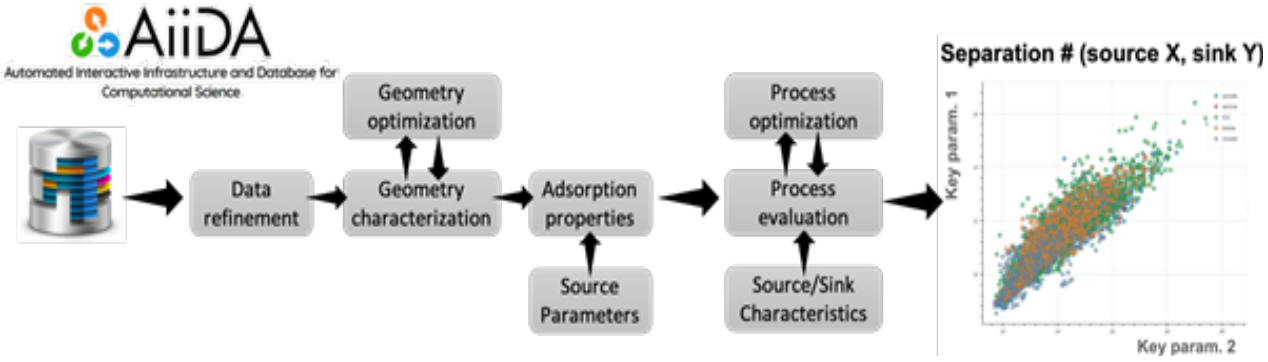


Figure 1. The screening platform: The workflow from importing a selected structure from an open-source database to screening and ranking its performance for a particular carbon capture application.

In the following section we provide an example of how the screening platform works for the particular case study of CO_2 capture from waste-to-energy power plants. The characteristics of such a flue gas stream are, among others, a CO_2 concentration of 13.6 % at 160 ºC temperature and 1 bar pressure. We also compare this to an active learning approach using PyePAL.

Case Study: Waste-to-Energy

In this article, we applied our TVSA process model to a particular binary mixture of CO_2 and N_2 (CO_2 concentration of 13.6 vol%) for CO_2 capture from waste-to-energy power plants. We selected 613 structures for screening and ranked their performance for a specific set of key performance indicators (e.g., specific work used for operating the separation process, CO_2 purity of the product stream, and working capacity). The TVSA process is illustrated in Figure 2 and consists of 4 steps: (i) adsorption, (ii) vacuum, (iii) heating under vacuum, and (iv) open cooling (and column pressurization).

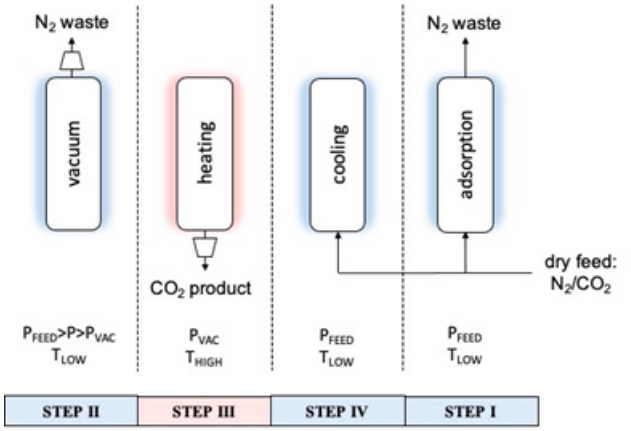


Figure 2. A 4-step TVSA process: (i) adsorption, (ii) vacuum, (iii) heating under vacuum, and (iv) open cooling.

Materials can be evaluated in our TVSA process model. For this evaluation the following assumptions were made:

- The feed stream downstream of the adsorption process is available at 20 ºC and 1 bar pressure.
- The temperature at steps (i), (ii), and (iv) is constant at 20 ºC. The temperature at step (iii) is constant at 120 ºC.
- Vacuum at steps (ii) and (iii) is applied at 0.05 bar. The rest of the steps operate at 1 bar pressure.
- The fluid phase is treated as an ideal gas.
- The multi-component adsorption equilibrium is described by the ideal adsorbed solution theory [10] and evaluated using the open-access python package pyIAST [11].
- A heat exchange fluid is used for heating and cooling and its temperature is homogeneous along the column.

As illustrated in the Figure 3, only a small number of the structures perform well for carbon capture from a waste-to-energy power plant.

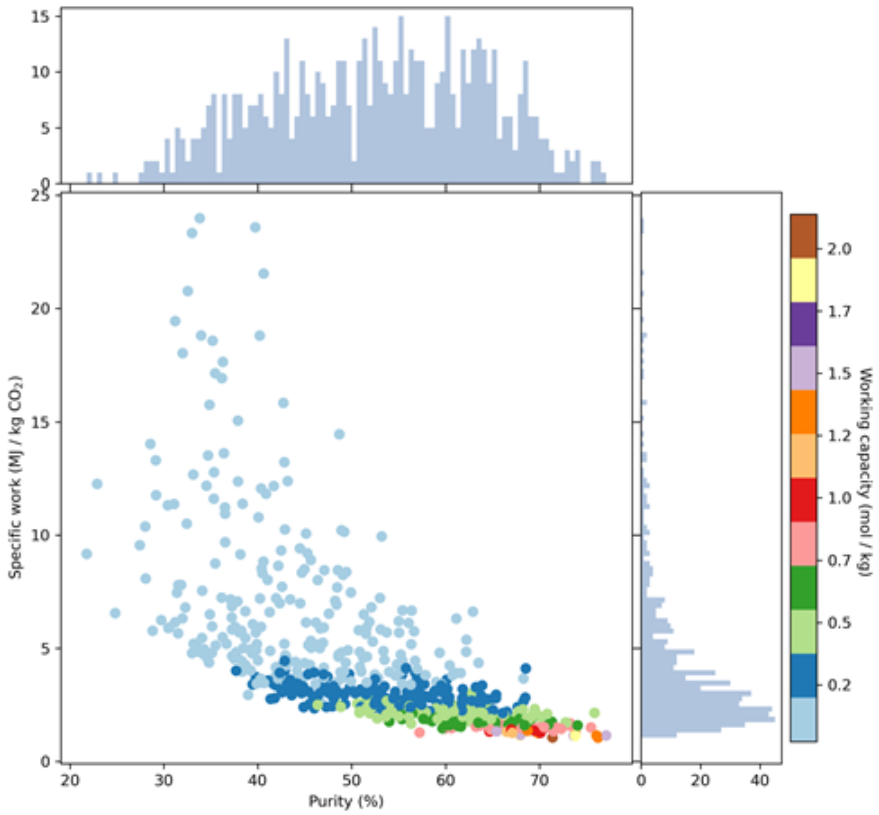


Figure 3. Specific work versus CO_2 purity and working capacity results from evaluating the performance of 613 structures in the TVSA process for CO_2 capture from waste-to-energy power plants. Histograms are shown for each axis.

Screening millions of *in silico* structures is computationally expensive. Evaluating the performance of each of the 613 structures in the TVSA process model took 3 hours (using a 2.7 GHz Dual-Core Intel Core i5 processor). An even more computationally expensive task is computing the physico-chemical (i.e., crystal density, crystal void fraction, heat capacity of the solid) and adsorptive properties (i.e., CO_2 and N_2 adsorption isotherms and heat of adsorption at different uptakes) of each material. Computing these properties for the 613 structures, which are used as inputs to the TVSA process model, took from 3 to 4 weeks (using an Intel Broadwell based cluster, Fidis [12]).

For the active learning approach, the design space is first initialized by a diverse set of materials. Material descriptors were taken from Seyed et al. [13] and the CoRE-MOF database [14]. These included the MOF-adapted Revised Auto Correlations (RACs), and geometric descriptors. The heat capacity for each material was also used as a descriptor. Three gaussian regressors (one for each KPI: specific work, working capacity, and purity) were trained on these initial materials and the KPIs predicted for each material. The gaussian regressors then predict KPIs for all other materials, also returning an uncertainty. Using PyePAL, these predictions and uncertainties are used to classify materials as Pareto-optimal, discarded, or remain unclassified. A new material is then 'sampled' which involves using the TVSA model to return accurate KPIs. The gaussian models are then retrained. This process is iterated until all points have been classified. At the end of this iterative process, all points have been classified, and we have a Pareto-optimal material set, a sampled set, and the trained models. The sampled set and Pareto-optimal set are seen in Figure 4. The grey points are the material KPIs as calculated by the TVSA process and remain unsampled as they are not deemed relevant in the training of the regressors. In our case, we have sampled 195 materials (seen as the green points) out of 613. The points identified as Pareto-optimal are shown as red stars.

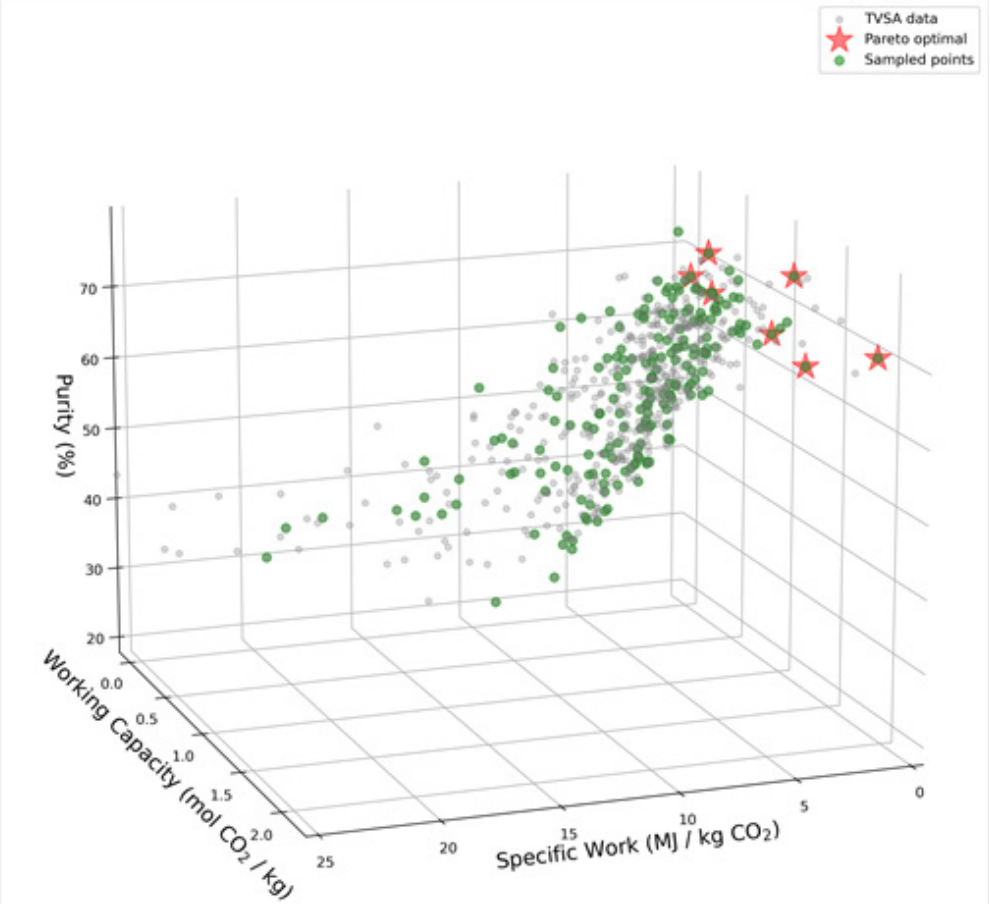


Figure 4. KPIs for each material as calculated by the TVSA process model are shown in grey. PyePAL samples points (green) as it builds a training set most relevant to Pareto-optimal materials (stars).

In terms of execution time, the brute-force technique where the TVSA model is used to calculate KPIs for all 613 materials takes 3 hours. Using PyePAL, only 195 materials are sampled which takes 1 hour. This time saving can be reduced further with a larger data set, as better predictive power from the regressors can be leveraged. As the number of MOFs within some databases venture into the trillions [15], time savings are increasingly important for MOF screening studies.

Future work with more materials will involve using the PyePAL algorithm to find important structure-property relationships. Specifically, we can discover what features are important to the regression models, as these give rise to good predictions for top-performers. In contrast, performing this analysis on a regression model trained without active learning would be biased towards the under-performing materials.

Acknowledgments

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. F.M. was supported by a James Watt Scholarship from the School of Engineering and Physical Sciences at Heriot-Watt University.

References

[1] S. Sircar, T. C. C. Golden, and M. B. B. Rao, “Activated carbon for gas separation and storage,” *Carbon N. Y.*, vol. 34, no. 1, pp. 1–12, Jan. 1996.

[2] R. Anderson and D. A. Gómez-Gualdrón, “Deep learning combined with IAST to screen thermodynamically feasible MOFs for adsorption-based separation of multiple binary mixtures,” *J. Chem. Phys*, vol. 154, p. 234102, 2021.

[3] K. Nagesh Pai, V. Prasad, and A. Rajendran, “Generalized, Adsorbent-Agnostic, Artificial Neural Network Framework for Rapid Simulation, Optimization, and Adsorbent Screening of Adsorption Processes,” *Cite This Ind. Eng. Chem. Res*, vol. 59, 2020.

[4] Z. Yao *et al.*, “Inverse design of nanoporous crystalline reticular materials with deep generative models,” *Nat. Mach. Intell. 2021 31*, vol. 3, no. 1, pp. 76–86, Jan. 2021.

[5] K. Maik Jablonka, G. Melpatti Jothiappan, S. Wang, B. Smit, and B. Yoo, “Bias free multiobjective active learning for materials design and discovery.”

[6] P. G. Boyd *et al.*, “Data-driven design of metal–organic frameworks for wet flue gas CO_2 capture,” *Nature*, vol. 576, no. 7786, pp. 253–256, Dec. 2019.

[7] L. Riboldi *et al.*, “Advanced methodology for screening of novel adsorption materials for cost-efficient CO_2 capture,” *SSRN Electron. J.*, Mar. 2021.

[8] L. Joss, M. Gazzani, M. Hefti, D. Marx, and M. Mazzotti, "Temperature Swing Adsorption for the Recovery of the Heavy Component: An Equilibrium-Based Shortcut Model," *Ind. Eng. Chem. Res.*, vol. 54, no. 11, pp. 3027–3038, Mar. 2015.

[9] M. Zuluaga, A. Krause, and M. Püschel, "ε-PAL: An Active Learning Approach to the Multi-Objective Optimization Problem," *J. Mach. Learn. Res.*, vol. 17, pp. 1–32, 2016.

[10] A. L. Myers and J. M. Prausnitz, "Thermodynamics of mixed-gas adsorption," *AIChE J.*, vol. 11, no. 1, pp. 121–127, Jan. 1965.

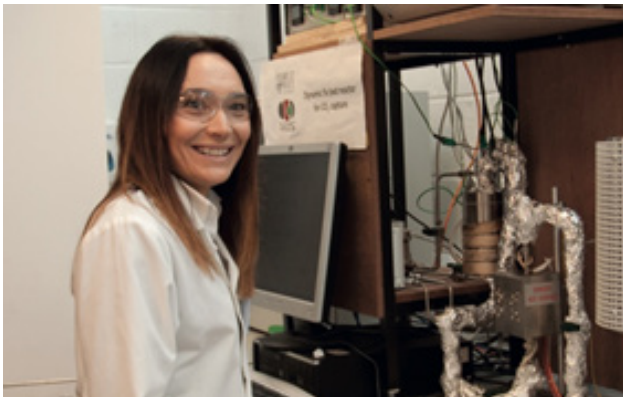
[11] C. M. Simon, B. Smit, and M. Haranczyk, "pyIAST: Ideal adsorbed solution theory (IAST) Python package," *Comput. Phys. Commun.*, vol. 200, pp. 364–380, Mar. 2016.

[12] "Fidis – Scientific IT and Application Support - EPFL." [Online]. Available: https://www.epfl.ch/research/facilities/scitas/hardware/fidis/. [Accessed: 15-Jul-2021].

[13] S. Mohamad Moosavi *et al.*, "Understanding the diversity of the metal-organic framework ecosystem."

[14] Y. G. Chung *et al.*, "Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal−Organic Framework Database: CoRE MOF 2019," 2019.

[15] S. Lee *et al.*, "Computational Screening of Trillions of Metal−Organic Frameworks for High-Performance Methane Storage," *ACS Appl. Mater. Interfaces*, vol. 13, 2021.



Prof Susana Garcia
Professor in Chemical Engineering
Associate Director in Carbon Capture and Storage,
Research Centre for Carbon Solutions (RCCS)

School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, EH14 4AS
United Kingdom
Phone: +44 (0) 131 451 8083
Email: s.garcia@hw.ac.uk

Susana García is Professor in Chemical Engineering and Deputy Head of the Institute of Mechanical, Process and Energy Engineering at Heriot-Watt University, which she joined in May 2014. She is also the Associate Director in Carbon Capture and Storage at the Research Centre for Carbon Solutions (RCCS), an interdisciplinary world leading engineering centre, inspiring and delivering innovation for the wider deployment of technologies needed to meet necessary carbon targets. She was the recipient of the Society of Spanish Researchers in the UK (SRUK/CERU) Emerging Talent Award 2020, in recognition of her outstanding professional trajectory, strong scientific outputs and mentoring and leadership qualities.

Her overarching research theme is the design and development of processes using advanced materials for specific applications relevant to energy efficiency and environmental sustainability in different sectors. She ambitions to change the paradigm on how novel processes based on advanced materials are developed through the integration of process engineering and basic science.