Matching pure and mixture isotherms using an invertible autoencoder

Ran, Y. A.; Calero, S.; Vlugt, T. J.H.; Dubbeldam, D.

DOI
10.1080/00268976.2025.2564899

Publication date
2025

Document Version
Final published version

Published in
Molecular Physics

Taylor & Francis
Taylor & Francis Group

Molecular Physics

An International Journal at the Interface Between Chemistry and Physics

ISSN: 0026-8976 (Print) 1362-3028 (Online) Journal homepage: www.tandfonline.com/journals/tmph20

Matching pure and mixture isotherms using an invertible autoencoder

Y. A. Ran, S. Calero, T. J. H. Vlugt & D. Dubbeldam

To cite this article: Y. A. Ran, S. Calero, T. J. H. Vlugt & D. Dubbeldam (2025) Matching pure and mixture isotherms using an invertible autoencoder, Molecular Physics, 123:21-22, e2564899, DOI: 10.1080/00268976.2025.2564899

To link to this article: https://doi.org/10.1080/00268976.2025.2564899

Published online: 07 Oct 2025.

MOLECULAR PHYSICS
2025, VOL. 123, NOS. 21–22, e2564899 (9 pages)
https://doi.org/10.1080/00268976.2025.2564899

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

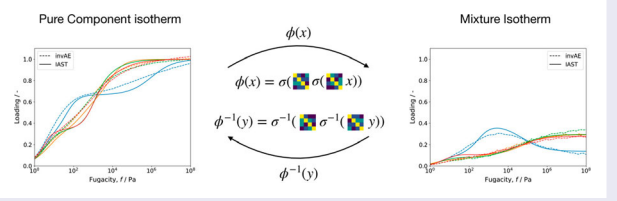
Matching pure and mixture isotherms using an invertible autoencoder

Y. A. Ran [a], S. Calero [b], T. J. H. Vlugt [c] and D. Dubbeldam [a]

[a]Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Amsterdam, The Netherlands; [b]Department of Applied Physics and Science Education, Eindhoven University of Technology, Eindhoven, The Netherlands; [c]Engineering Thermodynamics, Process & Energy Department, Faculty of Mechanical Engineering, Maritime and Materials Engineering, Delft University of Technology, Delft, The Netherlands

ABSTRACT

Porous materials such as zeolites and Metal-Organic Frameworks are widely used for molecular separations based on adsorption and enthalpy/entropy characteristics. Ideal adsorption solution theory (IAST) predicts mixture adsorption behaviour on the basis of pure component isotherms of adsorbents in porous media. Mixture data at all mole fractions are required for breakthrough simulations. The use of IAST avoids the expensive computations of mixtures with Monte Carlo methods. Matching outcomes from computational physics studies to experimentally measurable properties is the foundation of the materials design pipeline. Here, we report the regression of an Invertible Autoencoder (IAE) for the forward and backward mapping of pure and mixture isotherms. The invertible autoencoder is defined as a soft-invertible neural network, which can be used as mapping function. Pure component isotherms are modelled using a 3-site Langmuir-Freundlich model, with a broad range of equilibrium pressure and heterogeneity factors. A synthetic dataset is generated from pure component isotherms and mixture isotherms calculated with RUPTURA. The IAE predicts pure and mixture isotherms with high precision over a large fugacity range, for up to 6 components and 3-site isotherms. This work contributes to inverting the full design pipeline from physical gas separation to adsorbate design, enabling property-guided materials discovery.



ARTICLE HISTORY
Received 31 March 2025
Accepted 14 September 2025

KEYWORDS
RUPTURA; IAST; RASPA; adsorption

1. Introduction

Chemical separations based on heat cost approximately 20% of the world's energy [1], highlighting the need for more energy-efficient alternatives. Finding ways to separate chemicals at ambient temperatures and pressures is key to the global energy transition. Porous materials such as zeolites and metal-organic frameworks (MOFs) are indispensable for separations based on shape and affinity characteristics [2–4]. Nanoporous materials are used to great success for many separations, such as branched alkanes and paraffins [5, 6], but effective materials are yet to be discovered for other challenging mixtures, such as carbon capture from flue gases and perfluoroalkyl and polyfluoroalkyl substances [7–9].

Recent developments have led to the availability of databases with lots of nanoporous material structures, such as the CoRE MOF database [10]. The vastness of the available chemical space requires data-driven methodologies to effectively find materials with favourable adsorption properties [11, 12]. Adsorption properties, especially for mixtures, are essential for evaluating the ability of a candidate material to separate target components at industrially relevant conditions. Whereas single-component isotherms can be modelled with relatively straightforward adsorption models (e.g. Langmuir, Langmuir-Freundlich) or computed through molecular simulations [13], mixture adsorption introduces considerably more complexity [14–16]. Thermodynamically consistent models such as Ideal Adsorbed Solution Theory (IAST) or advanced simulations with RUPTURA are used to predict mixture loadings at process conditions. However, running simulations increases the

CONTACT D. Dubbeldam d.dubbeldam@uva.nl Van 't Hoff Institute for Molecular Sciences, University of Amsterdam, Science Park 904, Amsterdam 1098 XH, The Netherlands

computational cost for large-scale screening. In many practical scenarios, mixture isotherms are more readily measured than pure component isotherms, or a target mixture behaviour is defined in a material design workflow, motivating an invertible formulation of IAST.

Machine learning models assist in predicting adsorption properties from descriptors of chemical composition and pore structure, whereas inverse design tasks identify materials that meet desired property targets [17–23]. It is important to note that an ongoing challenge remains: bridging the gap between accurately describing adsorption isotherms for the entire pressure range spanning the Henry regime at low loading through to saturation at high fugacity, and enabling a truly invertible framework that can transition seamlessly between adsorptive properties and the structural/chemical design space.

Here, we report the regression of an Invertible Autoencoder (IAE) for forward and backward mapping of both pure and mixture adsorption isotherms. This is the first part of a computational pipeline that we are developing that in the future hopefully will be able to extract mixture and pure component isotherms from (experimental) breakthrough data. For now, we focus on the invertibility (back and forth) of pure component and mixture data. In generating our training data, we build on a 3-site Langmuir-Freundlich isotherm to model pure-component adsorption for a broad range of fugacities and heterogeneity factors. We compute synthetic isotherms using RUPTURA [16], creating a large and diverse dataset (Appendix 1). We demonstrate that the proposed IAE achieves high accuracy in both the low-pressure Henry regime and high-pressure saturation region for up to six-component mixtures. By learning invertible relationships between system parameters – fugacity, composition and heterogeneity – and adsorption loadings this work contributes to the design process by enabling backward mapping of mixture isotherms onto pure isotherms.

2. Methods

2.1. Isotherm models

Adsorption behaviour can be accurately represented through the usage of functional forms that describe adsorption. Sharma et al. have provided an excellent review on isotherm and mixture models accompanying the RUPTURA code [16]. In the simplest case, adsorption is modelled as a monolayer adsorbing onto open sites (s_{open}) on a surface, as proposed by Langmuir [24]. The adsorption isotherm $q(f)$ describes the loading as a function of the fugacity f. The Langmuir model relates the adsorption rate $\dot{q}_{\text{ads}} = kfs_{\text{open}}$, and the desorption rate $\dot{q}_{\text{des}} = k'q$ for total sites $s_{\text{total}} = s_{\text{open}} + q$. At equilibrium, the adsorption and desorption rates are equal, yielding

$$q(f) = q_{\text{sat}} \frac{bf}{1 + bf} \tag{1}$$

for the rate constant $b = \frac{k}{k'}$, and saturation loading equal to the total number of sites ($q_{\text{sat}} = s_{\text{total}}$). One feature of the Langmuir equation is that at low fugacity the partial derivative $\lim_{f \to 0} \mathrm{d}q/\mathrm{d}f$ equals the rate constant, also named the Henry constant b. The Langmuir equation can be solved for 50% adsorption $q(f_{50}) = \frac{1}{2} q_{\text{sat}}$, which yields $f_{50} = \frac{1}{b}$. When plotting the Langmuir equation on the semi-logarithmic scale an inflection point can be found at exactly f_{50}, as the derivative has a maximum at that point. A multisite adsorption isotherm can easily be obtained by a summation over multiple sites

$$q(f) = \sum_{i=0}^{n_{\text{sites}}} q_{i,\text{sat}} \frac{b_i f}{1 + b_i f} \tag{2}$$

The Freundlich isotherm accounts for the heterogeneity of the different modes of adsorption by posing that the adsorption rate, $\dot{q}_{\text{ads}}$ does not increase linearly with fugacity, but has a power law scaling determined by factor ν. Incorporating this into the multisite Langmuir equation yields the multisite Langmuir-Freundlich equation. The Langmuir-Freundlich equation is the most common model describing isotherms in nanoporous materials.

$$q(f) = \sum_{i=0}^{n_{\text{sites}}} q_{i,\text{sat}} \frac{b_i f^{\nu}}{1 + b_i f^{\nu}} \tag{3}$$

While this is an accurate description of the behaviour in the inflection points and high fugacity ranges, this model no longer yields a correct theoretical limit in the Henry regime, as the derivative $\lim_{f \to 0} \mathrm{d}q/\mathrm{d}f$ yields ∞ for $\nu < 1$ and 0 for $\nu > 1$.

2.2. Ideal adsorbed solution theory (IAST)

The aforementioned isotherms assume the adsorption of a single component, but for many practical applications it is necessary to evaluate the competitive adsorption of two or more components. Theoretical descriptions of mixture isotherms, i.e. competition at a binding site, have appeared for simple systems, such as two Langmuir components with identical q_{sat} parameters. For more complex mixtures and adsorption processes an alternative process has been developed, named Ideal-Adsorbed Solution Theory (IAST) [25–28].

IAST provides relatively simple calculations for estimating multicomponent adsorption from single-com

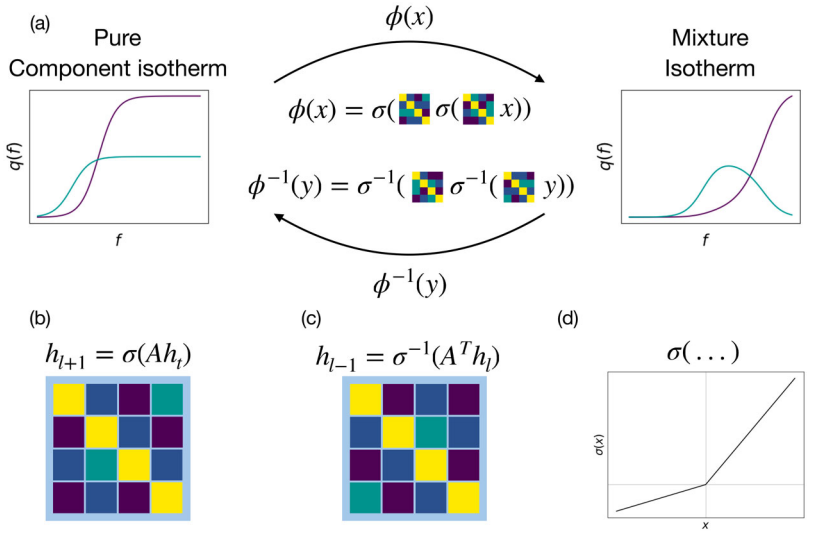


Figure 1. Schematic overview of the structure of the IAE. (a) A mapping from the pure isotherms x to the mixture isotherms y, using neural network ϕ and the reverse mapping ϕ^{-1} occur as a series of matrix-vector products and non-linear transformations. The forward (b) and reverse (c) operation is given by a matrix-vector product with weight matrix A and A^T respectively. (d) A non-linear transformation is done with the leaky RELU operation, doubling all positive and halving negative values.

ponent isotherms and is widely applicable across diverse gas mixtures. However, it assumes ideal solution beha viour in the adsorbed phase and neglects possible competitive effects and surface heterogeneity, which can lead to inaccuracies in systems with strong adsorbate-adsorbent interactions. While more complex mixture adsorption may occur, captured for example by non-ideal activity coefficients [29] or size-dependent isotherms [30], this work restricts to adsorption that is modelled by IAST. In modelling the adsorption of fluid mixtures within porous materials, the reduced grand potential, ψ, is used. It is defined as

$$\psi_i(f_i^*) = \int_0^{f^*} \frac{q_i^*(f)}{f}\,\mathrm{d}f, \tag{4}$$

for pure component adsorption q^* and pure component fugacity f^*, and hence can be easily computed from the adsorption isotherm. When applying IAST, the equilibrium criterion is that the fugacity of each component in the bulk phase equals the corresponding fugacity in the adsorbed phase. Expressing this in terms of ψ shows that all components share the same reduced grand potential at equilibrium. The IAST balance for an ideal solution is

$$y_i p_T = x_i p_i^*(\psi) \tag{5}$$

where y_i is the bulk-phase mole fraction, x_i the adsorbed-phase mole fraction, p_T the total pressure, and $p_i^*(\psi)$ the pure component pressure corresponding to the same ψ. This equation can be solved by imposing that the sum of all x_i equals unity.

2.3. Invertible neural net

The invertible autoencoder is used as defined in [31] and is schematically represented in Figure 1. The IAE is a neural network $\phi : R^n \to R^n$, which has an inverse $\phi^{-1} : R^n \to R^n$. The network is optimised to, but not equal to, $\phi\phi^{-1} = 1$.

The pure and mixture isotherms are calculated as a set of isotherms for six components as described in Appendix 1. The training data therefore consists of n_{train} sets of pure and mixture isotherms $q_{\text{pure}}(\mathrm{c}, f)$ and $q_{\text{mix}}(\mathrm{c}, f)$, for component c and fugacity f.

The fully connected linear layer has a forward defined by $h_{l+1} = Ah_l$ and an inverse defined as $h_{l-1} = A^T h_l$. This is not an exact inverse as A^T only equals A^{-1} for orthonormal matrices. Restricting the learning process to just orthonormal matrices is not practical for training and restricts the neural net by disallowing dimensional reduction and expansion. By training reversibility by mapping the reverse with A^T the network will achieve a soft reversibility $AA^T \approx \mathbf{1}$.

A leaky ReLU is a non-linear activation function that still preserves reversibility:

$$h_{l+1}^i = \begin{cases} \frac{1}{a} h_l^i, & h_l^i < 0 \\ a h_l^i, & h_l^i \geq 0 \end{cases} \tag{6}$$

where h_l^i is the i-th value of the input vector h_l. The inverse is found by replacing a with a^{-1} (here $a = 2$). Weight parameters are optimised using the Adam optimiser [32], using gradient clipping, L2 weight decay and k-folding. Hyperparameters are listed in Appendix 2. The

optimal joint likelihood is defined as the geometric mean between the likelihood of finding the forward mapping and the reverse mapping

$$\mathcal{L}(\hat{X},\hat{Y}\,|\,X,Y,\theta) = \sqrt{\mathcal{L}(\hat{X}\,|\,X,\theta)\mathcal{L}(\hat{Y}\,|\,Y,\theta)} \quad (7)$$

Where the prediction $\hat{X} = \phi^{-1}(Y)$ and $\hat{Y} = \phi(X)$. Optimizing the log-likelihood of this function is equal to minimizing a mean squared error

$$L(\hat{X},\hat{Y}\,|\,X,Y,\theta) = \sum_{i=0}^{n_{train}} (\phi(x_i) - y_i)^2 + (\phi^{-1}(y_i) - x_i)^2 \quad (8)$$

3. Results

Figure 2 illustrates the mean squared error (MSE) of the IAE on the test set as a function of the number of training samples. For fewer than 6×10^4 training samples the MSE is relatively high, indicating that only limited mapping information is captured with limited training data. In the range of samples $[6 \times 10^4, 10^6]$, the MSE decreases according to the expected trend $1/\sqrt{n_{\text{train}}}$, indicating that increased training data improve accuracy. Beyond 10^6 samples, the MSE does not further decrease following the expected $1/\sqrt{n_{\text{train}}}$ trend, showing diminishing returns for increased costs. Any training size within the interval $[6 \times 10^4, 10^6]$ is sufficient to achieve accurate representation, with MSE values reaching 3×10^{-6} per sample. The training is considered converged when the validation loss shows no improvement larger than $\varepsilon = 10^{-3}$ for 5 epochs, as shown in the training curves in Figure A1. This results in a model that accurately predicts isotherms, as shown in the parity plots in Figure 3.

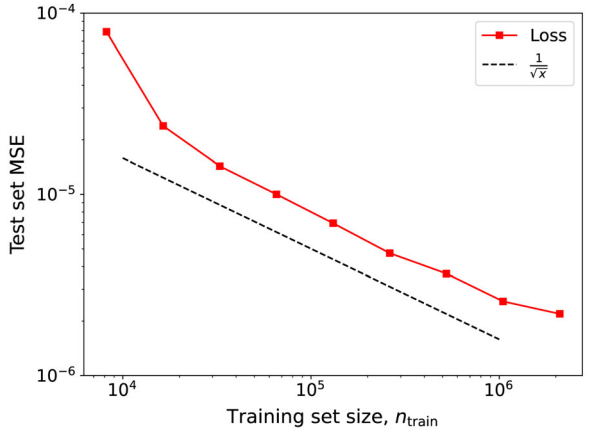


Figure 2. Mean Squared Error (MSE) loss of the model with respect to the size of the training set.

3.1. Statistical significance

Although the random sample from the test set provides an indication of performance, a more rigorous approach requires analysing the coefficient of determination R^2 at each fugacity over the test set. These overall R^2 values on the test set appear in Figure 4. In our dataset, the highest density of inflection points occurs around 10^4 Pa, explaining the lower R^2 values in that range. We also observe that the model performs best for 3-5 component systems and is not as effective for 2-component systems.

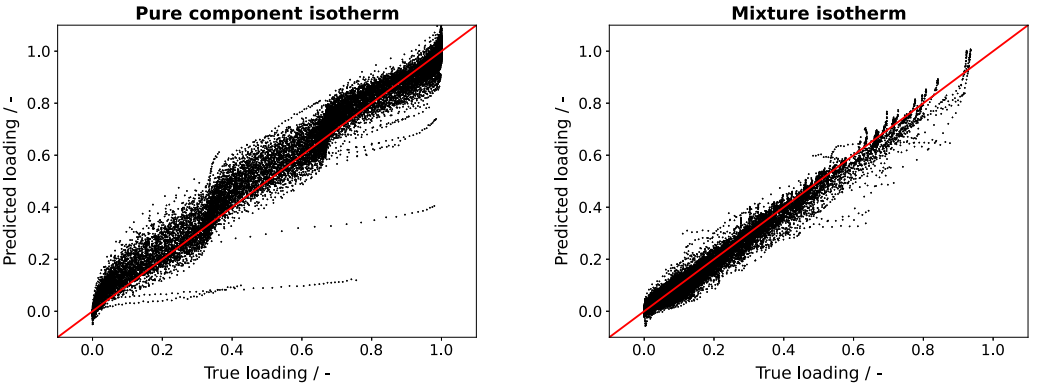
3.2. Mapping of isotherms

In Figure 5 we show representative pure and mixture isotherm for a system with four compounds. This sample has been randomly selected from the test set. The pure isotherms (Figure 5(a)) show sigmoidal adsorption behaviour in semi-logarithmic scale, with up to 3 inflection points. From these pure adsorption isotherms, a mixture isotherm is calculated using IAST and shown in Figure 5(b). Here, the compound that adsorbs best at lower pressures (green) dominates adsorption at lower fugacities in the mixture. The pure isotherms obtained by the IAE mapping of the mixture isotherm closely match the true pure isotherms. At all points, the pure adsorption isotherm is reproduced to within a small error. However, the model does not fully capture all inflections and their locations for all components. When mapping the pure isotherm to the mixture isotherm, shown in Figure 5(b), the model generally performs well for the given sample.

Figure 6 shows the isotherms on a log-log scale, highlighting the low loading regimes. The mappings of both pure to mixture and mixture to pure agree well in the Henry regime. However, for the mapping of the pure to mixture isotherms, an underestimation of a factor of two in loading is observed in the dilute region.

3.3. Varying number of components

By setting the adsorption isotherm of one component to zeros a component is knocked out and the same IAE neural net is applied to a different number of components. Figure 7 shows mappings of random samples from the test set with 2–6 components. Overall, the model performs best for a larger number of components, as seen for the 5- and 6-component system. In contrast, for two component isotherms, the pure component isotherm shows an overestimated Henry coefficient, which shifts the isotherm to lower pressures. Conversely, the mapping from pure to mixture the isotherms are shifted in the opposite direction. This behaviour likely stems from the fact that less information is captured in the isotherms,



(a) (b)

Figure 3. Parity plots showing the predictions of the model with respect to the actual values of the model, for (a) pure component isotherms, predicted from the mixture isotherms and (b) vice versa. Parity plots are shown for 100 pure component and mixture isotherms randomly selected from the test set.

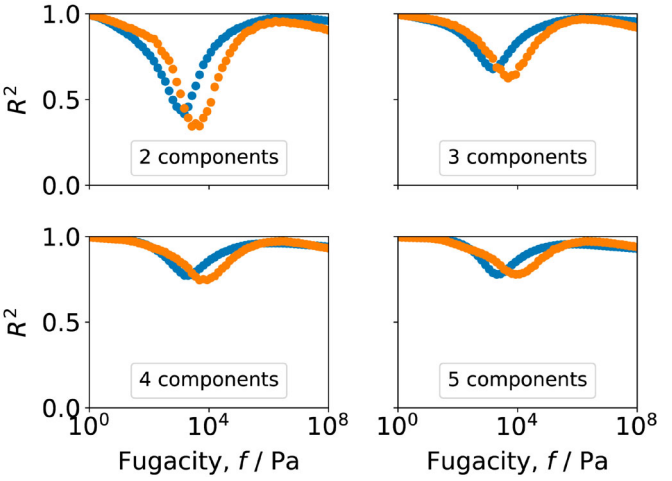
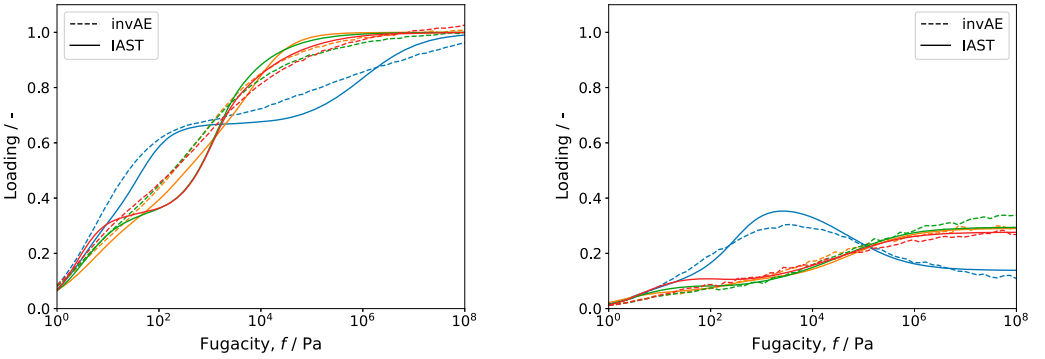


Figure 4. Statistical significance of the model R^2 computed per pressure point for mapping mixture to pure (blue) and pure to mixture (orange) isotherms.



(a) (b)

Figure 5. A random sample of pure (a) and mixture (b) isotherms computed by RUPTURA (solid) and the invertible autoencoder (dashed). Langmuir-Freundlich parameters for the isotherms in this figure is shown in Table A6.

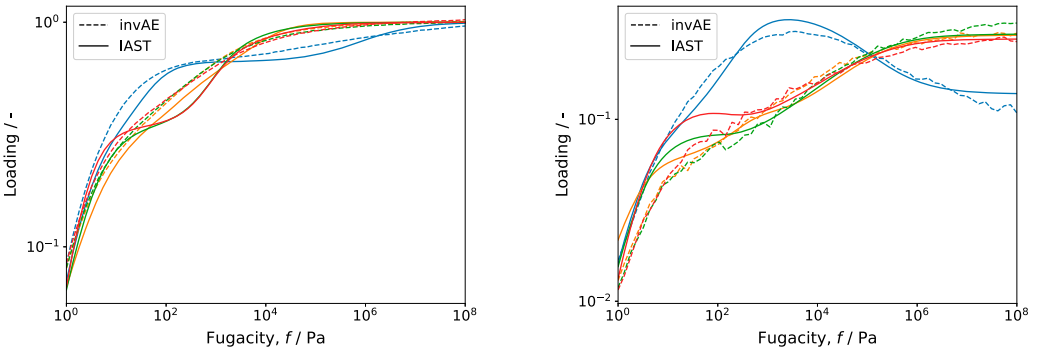


Figure 6. A random sample of pure (a) and mixture (b) isotherms on a log scale computed by RUPTURA (solid) and the invertible autoencoder (dashed). Langmuir-Freundlich parameters for the isotherms in this figure is shown in Table A6.

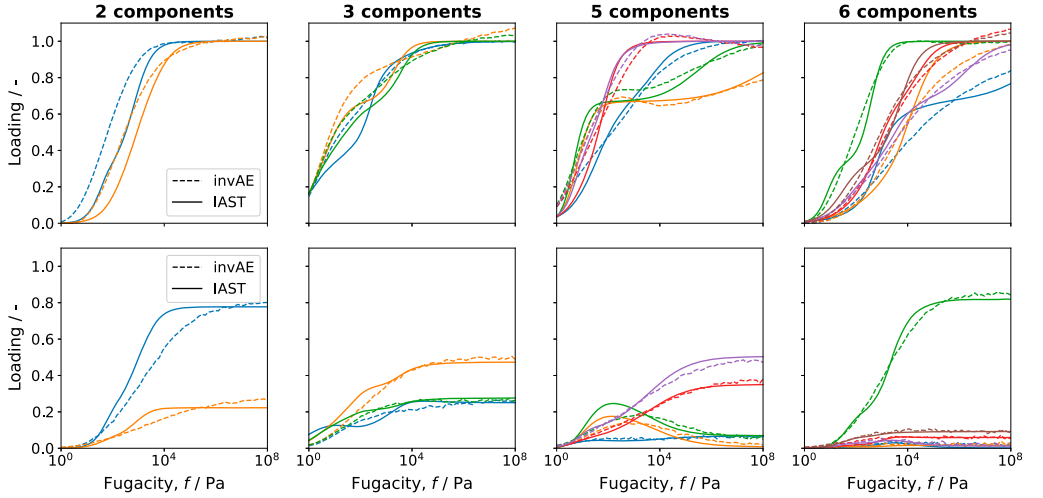


Figure 7. Random samples of pure (top) and mixture (bottom) isotherms for varying component numbers computed by RUPTURA (solid) and the invertible autoencoder (dashed). Langmuir-Freundlich parameters for the isotherms in this figure is shown in Table A4, A5, A7 and A8.

making it harder to reconstruct the map from it. The data imbalance in the training set might also have an influence, as shown in the prevalence listed in Table A1, however, this does not seem to affect the 6-component isotherms.

4. Conclusions

In this work, we have successfully implemented a mapping function for the reversible conversion of pure-component and mixture loading isotherms. Now, mixture isotherms can conveniently be mapped back onto pure isotherms, possibly bypassing the need for single component experiments. The mapping is defined as a lightweight neural network modelled with soft invertible layers, allowing up to six-component isotherms to be mapped over a wide range of physical parameters (q_{ads}, b and ν). Overall, the invertible autoencoder approach captures the general shape and magnitude of both pure and mixture adsorption isotherms across a wide range of pressures. It accurately reproduces the dominant adsorption behaviour of the best adsorbing component at lower fugacities and shows good agreement in the Henry regime. Although it struggles with precisely locating inflection points in certain cases, the overall results are encouraging for accurately describing adsorption in multi-component isotherms. Building on this model a full reversible workflow can be envisioned, mapping breakthrough measurements directly onto material properties.

Acknowledgments

The authors thank Microsoft and Max Welling for supporting this project and discussion.

Data availability statement

The data sets generated and analysed during the current study are available in the FigShare repository of the University of Amsterdam at https://doi.org/10.21942/uva.30148591. The source code used for dataset creation and learning protocols is openly available on GitHub at https://github.com/YouriRan/invRUPTURA.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Y. A. Ran http://orcid.org/0009-0001-5065-0314
S. Calero http://orcid.org/0000-0001-9535-057X
T. J. H. Vlugt http://orcid.org/0000-0003-3059-8712
D. Dubbeldam http://orcid.org/0000-0002-4382-1509

References

[1] D.S. Sholl and R.P. Lively, Nature **532**, 435–437 (2016). doi:10.1038/532435a
[2] M.E. Davis, Nature **417**, 813–821 (2002). doi:10.1038/nature00785
[3] Y.S. Bae and R.Q. Snurr, Angewandte Chemie Int. Edition **50**, 11586–11596 (2011). doi:10.1002/anie.v50.49
[4] J.R. Li, R.J. Kuppler and H.C. Zhou, Chem. Soc. Rev. **38**, 1477–1504 (2009). doi:10.1039/b802426j
[5] D. Dubbeldam, C.J. Galvin, K.S. Walton, D.E. Ellis and R.Q. Snurr, J. Am. Chem. Soc. **130**, 10884–10885 (2008). doi:10.1021/ja804039c
[6] Z.R. Herm, B.M. Wiers, J.A. Mason, J.M. van Baten, M.R. Hudson, P. Zajdel, C.M. Brown, N. Masciocchi, R. Krishna and J.R. Long, Science **340**, 960–964 (2013). doi:10.1126/science.1234071
[7] N. Gargiulo, F. Pepe and D. Caputo, J. Nanosci. Nanotechnol. **14**, 1811–1822 (2014). doi:10.1166/jnn.2014.8893
[8] N. Jiang, R. Shang, S.G.J. Heijman and L.C. Rietveld, Water. Res. **144**, 145–161 (2018). doi:10.1016/j.watres.2018.07.017
[9] T. Ghanbari, F. Abnisa and W.M.A. Wan Daud, Sci. Total Environ. **707**,135090(2020). doi:10.1016/j.scitotenv.2019.135090
[10] Y.G. Chung, J. Camp, M. Haranczyk, B.J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O.K. Farha, D.S. Sholl and R.Q. Snurr, Chem. Mater. **26**, 6185–6192 (2014). doi:10.1021/cm502594j
[11] E. Braun, A.F. Zurhelle, W. Thijssen, S.K. Schnell, L.C. Lin, J. Kim, J.A. Thompson and B. Smit, Mol. Syst. Des. Eng. **1**, 175–188 (2016). doi:10.1039/C6ME00043F
[12] K.M. Jablonka, D. Ongari, S.M. Moosavi and B. Smit, Chem. Rev. **120**, 8066–8129 (2020). doi:10.1021/acs.chemrev.0c00004
[13] P.R. Van Tassel, H.T. Davis and A.V. McCormick, J. Chem. Phys. **98**, 8919–8928 (1993). doi:10.1063/1.464451
[14] T.R.C. Van Assche, G.V. Baron and J.F.M. Denayer, Adsorption **30**, 351–361 (2024). doi:10.1007/s10450-024-00437-y
[15] Y.A. Ran, S. Sharma, S.R.G. Balestra, Z. Li, S. Calero, T.J.H. Vlugt, R.Q. Snurr and D. Dubbeldam, J. Chem. Phys. **161**,114106(2024). doi:10.1063/5.0226249
[16] S. Sharma, S.R.G. Balestra, R. Baur, U. Agarwal, E. Zuidema, M.S. Rigutto, S. Calero, T.J.H. Vlugt and D. Dubbeldam, Mol. Simul. **49**, 893–953 (2023). doi:10.1080/08927022.2023.2202757
[17] B. Kim, S. Lee and J. Kim, Sci. Adv. **6**, eaax9324 (2020). doi:10.1126/sciadv.aax9324
[18] S.M. Moosavi, K.M. Jablonka and B. Smit, J. Am. Chem. Soc. **142**, 20273–20287 (2020). doi:10.1021/jacs.0c09105
[19] Z. Wang, T. Zhou and K. Sundmacher, Chem. Eng. J. **444**,136651(2022). doi:10.1016/j.cej.2022.136651
[20] Z. Wang, Y. Zhou, T. Zhou and K. Sundmacher, Comput. Chem. Eng. **160**,107739(2022). doi:10.1016/j.compchemeng.2022.107739
[21] Y.H. Cheng, I.T. Sung, C.M. Hsieh and L.C. Lin, J. Taiwan Inst. Chem. Eng. **165**,105728(2024). doi:10.1016/j.jtice.2024.105728
[22] I.T. Sung and L.C. Lin, Phys. Chem. C **127**, 13886–13899 (2023). doi:10.1021/acs.jpcc.3c02452
[23] D. Ongari, L. Talirz, K.M. Jablonka, D.W. Siderius and B. Smit, Chem. Eng. Data **67**, 1743–1756 (2022). doi:10.1021/acs.jced.1c00958
[24] I. Langmuir, J. Am. Chem. Soc. **40**, 1361–1403 (1918). doi:10.1021/ja02242a004
[25] A.L. Myers and J.M. Prausnitz, AIChE J. **11**, 121–127 (1965). doi:10.1002/aic.v11:1
[26] K.S. Walton and D.S. Sholl, AIChE J. **61**, 2757–2762 (2015). doi:10.1002/aic.v61.9
[27] M. Murthi and R.Q. Snurr, Langmuir **20**, 2489–2497 (2004). doi:10.1021/la035556p
[28] A.L. Myers and P.A. Monson, Adsorption **20**, 591–622 (2014). doi:10.1007/s10450-014-9604-1
[29] C.C. Brunchi, P. Englebienne, H.J.M. Kramer, S.K. Schnell and T.J.H. Vlugt, Mol. Simul. **41**, 1234–1244 (2015). doi:10.1080/08927022.2014.972394
[30] S. Sharma, M.S. Rigutto, R. Baur, U. Agarwal, E. Zuidema, S.R.G. Balestra, S. Calero, D. Dubbeldam and T.J.H. Vlugt, Mol. Phys. **121**, e2183721 (2023). doi:10.1080/00268976.2023.2183721
[31] Y. Teng and A. Choromanska, Computation **7** (2), 20 (2019). doi:10.3390/computation7020020
[32] D.P. Kingma and J. Ba, in *3rd International Conference for Learning Representations* (San Diego, 2015). doi:10.48550/arXiv.1412.6980

Appendices

Appendix 1. Synthetic data

For creating the pure and mixture isotherms we use the RUPTURA package [16]. To create pure component adsorption isotherms, the multisite Langmuir-Freundlich isotherm is evaluated on 64 fugacity points evenly spaced on a logscale from $1 - 10^8$ Pa. Langmuir-Freundlich parameters (q_s^c, b_s^c, ν_s^c) are chosen pe r component (c) and binding site (s) such that they represent a broad range of possible binding curves within the physical range. q_s^c is chosen as $(n_{\text{sites}})^{-1}$, as this normalisation leads to the most stable training data. A set of (q_s^c, b_s^c, ν_s^c) is chosen for all six components and 3 adsorption sites, yielding 54 parameters. From this, the pure and mixture isotherms are computed and discretizations are stored as pytorch tensors of shape (6, 64). The parameters are sampled such that most inflection points are captured in the given fugacity region and the adsorption is competitive.

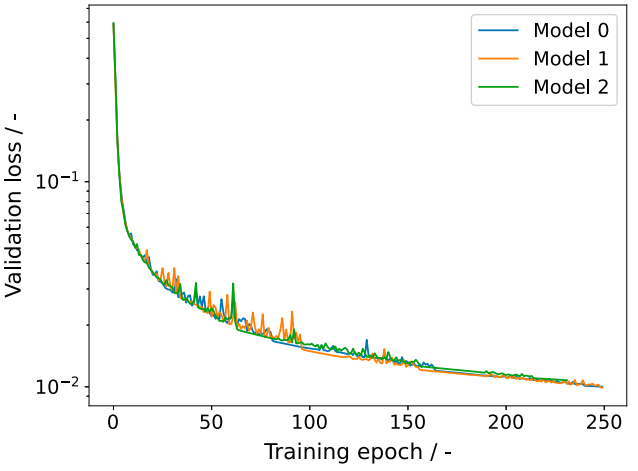


Figure A1. Training curves for the optimisation of 3 models of the invertible autoencoder, using different k-folds of the data.

Table A1. Fraction ϕ of numbers of isotherms N in the training data set given by a knockout probability of $p = 0.4$.

N	0	1	2	3	4	5	6
ϕ	0.004	0.037	0.138	0.276	0.311	0.187	0.047

The equilibrium constant b_s^c is chosen in such a way that the inflection points mainly fall within the evaluated pressure range. The logarithm of the equilibrium constant $\log(b_s)$ is sampled from a uniform distribution $\mathcal{U}(-6, 0.5)$. The equilibrium binding constant of components per site is often seen to be similar between components compared in experiment and therefore $\log(b_s^c) = (1 + \delta)\log(b_s)$ and $\delta \sim \mathcal{U}(-\epsilon, \epsilon)$, such that small relative variations per component are generated at a site.

The parameter ν can be regarded as the conformational entropy at a site, and represents itself in the steepness of the isotherm curve. It also shifts the inflection point, such that the inflection point $\log(f_{50}) = \frac{1}{\nu}\log(b)$. By sampling $\nu \sim \mathcal{U}(0.4, 2.0)$ a physical range is represented and the inflection points are uniformly distributed along the pressure range.

In order to create a dataset with a varied number of isotherms of one to six components, during data creation, the components are *knocked out*. This is done by setting the b parameters of the knocked out isotherms to an arbitrarily small positive number, 10^{-30} Pa^{-1} here, setting both the pure and mixture isotherm to effectively zero for both the pure and mixture. Throughout the dataset this is done with a probability of $p = 0.4$, such that a set of isotherms can contain multiple knocked out components, yielding data for 1 to 6 component isotherms, with probabilities governed by the prevalence factions given in Table A1.

Appendix 2. Invertible autoencoder

We employed an IAE architecture comprising eight invertible layers, each represented by a matrix whose dimensions are summarised in Table A2.

We employed the hyperparameters summarised in Table A3 for training. A large training set was used for the evaluations reported in Figures 5–4, consistent with the relationship illustrated in Figure 2, which shows the mean squared error (MSE) as a function of the training set size. Although 3-fold cross-validation was performed, the models converged to effectively identical solutions; hence, cross-model errors were negligible and are not reported. Training progress is tracked through the average training loss of the individual models shown in Figure A1.

Table A2. IAE weight matrix architecture.

W_0	W_1	$W_2 - W_5$	W_6	W_7
384 × 128	128 × 64	64 × 64	64 × 128	128 × 384

Table A3. Hyperparameters used during training of the IAE.

ntrain	nval	lr	weight decay	k-folds
2^{21}	2^{18}	10^{-4}	10^{-5}	3

Table A4. Randomly sampled Langmuir-Freundlich parameters for a 2-component isotherm.

Component	Site	q_{sat}	b	ν
1	1	1/3	1.56×10^{-4}	1.30
	2	1/3	3.77×10^{-3}	1.10
	3	1/3	9.07×10^{-5}	1.09
2	1	1/3	1.40×10^{-4}	1.28
	2	1/3	3.71×10^{-3}	1.62
	3	1/3	1.23×10^{-4}	1.44

Table A5. Randomly sampled Langmuir-Freundlich parameters for a 3-component isotherm.

Component	Site	q_{sat}	b	ν
1	1	1/3	3.31×10^{-5}	1.36
	2	1/3	7.79×10^{-1}	1.38
	3	1/3	5.42×10^{-2}	1.30
2	1	1/3	3.18×10^{-5}	1.26
	2	1/3	7.82×10^{-1}	1.30
	3	1/3	5.29×10^{-2}	0.91
3	1	1/3	4.41×10^{-5}	1.83
	2	1/3	7.74×10^{-1}	0.89
	3	1/3	5.54×10^{-2}	0.47

Table A6. Randomly sampled Langmuir-Freundlich parameters for a 4-component isotherm.

Component	Site	q_{sat}	b	ν
1	1	1/3	4.24×10^{-3}	0.93
	2	1/3	2.62×10^{-5}	1.18
	3	1/3	2.41×10^{-1}	0.91
2	1	1/3	3.34×10^{-3}	0.70
	2	1/3	2.29×10^{-5}	1.52
	3	1/3	2.34×10^{-1}	1.21
3	1	1/3	3.90×10^{-3}	0.63
	2	1/3	2.01×10^{-5}	1.57
	3	1/3	2.42×10^{-1}	1.59
4	1	1/3	4.25×10^{-3}	1.45
	2	1/3	2.86×10^{-5}	0.76
	3	1/3	2.61×10^{-1}	1.27

Table A7. Randomly sampled Langmuir-Freundlich parameters for a 5-component isotherm.

Component	Site	q_{sat}	b	ν
1	1	1/3	4.70×10^{-2}	1.60
	2	1/3	4.93×10^{-4}	0.41
	3	1/3	6.99×10^{-2}	1.42
2	1	1/3	4.08×10^{-2}	1.83
	2	1/3	4.81×10^{-4}	0.61
	3	1/3	7.21×10^{-2}	1.94
3	1	1/3	3.92×10^{-2}	0.89
	2	1/3	5.67×10^{-4}	1.69
	3	1/3	7.38×10^{-2}	0.98
4	1	1/3	4.29×10^{-2}	1.81
	2	1/3	6.93×10^{-4}	1.68
	3	1/3	8.51×10^{-2}	0.69
5	1	1/3	3.93×10^{-2}	0.87
	2	1/3	7.02×10^{-4}	0.84
	3	1/3	7.05×10^{-2}	0.81

Table A8. Randomly sampled Langmuir-Freundlich parameters for a 6-component isotherm.

Component	Site	q_{sat}	b	ν
1	1	1/3	1.49×10^{-5}	1.23
	2	1/3	9.16×10^{-5}	0.89
	3	1/3	3.23×10^{-2}	0.56
2	1	1/3	2.52×10^{-5}	1.77
	2	1/3	9.28×10^{-5}	1.58
	3	1/3	2.51×10^{-2}	1.81
3	1	1/3	1.52×10^{-5}	1.09
	2	1/3	1.32×10^{-4}	1.37
	3	1/3	2.55×10^{-2}	0.72
4	1	1/3	3.94×10^{-5}	0.70
	2	1/3	1.51×10^{-4}	1.10
	3	1/3	3.15×10^{-2}	0.81
5	1	1/3	2.33×10^{-5}	1.26
	2	1/3	1.73×10^{-4}	1.00
	3	1/3	2.97×10^{-2}	1.06
6	1	1/3	1.78×10^{-5}	1.69
	2	1/3	1.74×10^{-4}	0.42
	3	1/3	2.43×10^{-2}	0.57

Appendix 3. Figure metadata

In this section the randomly sampled parameters to produce the pure component and mixture component isotherms are listed.