MACEIÓ – ALAGOAS
EBA 15
15° ENCONTRO BRASILEIRO SOBRE ADSORÇÃO
15th Brazilian Meeting about Adsorption

Energy histograms as features for machine learning to predict adsorption in zeolite

Maxwell T. Monte[a], Maria S. P. Aquino[a], José C. A. Oliveira [a], Andréa S. Pereira[a], Pedro F. G. Silvino[a], Sebastião M. P. Lucena[a]

[a]*Laboratory of Modeling and 3D Visualization, Department of Chemical Engineering, Universidade Federal do Ceará, Campus do Pici, bl. 709, 60455-760, Fortaleza, CE, Brazil*

Abstract

This study has the objective of search and study materials with well-suited properties in material science and engineering on the subject of adsorption, focusing on the storage of new combustibles like hydrogen. As a result of the compatibility and the needs of this task zeolite structures were chosen to be the material of study because of their high adsorption capacity, chemical and thermal stability. The research of the best structures among the zeolites was made by the use of machine learning, substituting the need of experimental tests, using as algorithm parameters data of energy histograms created by calculating the potential energies of the intermolecular interactions of a hydrogen molecule in 1e5 random points inside the zeolite structures. To make this research 42 zeolite structures of IZA database were chosen randomly to make the energy histograms and feed the machine learning algorithm, resulting in two trained models with great capacity to evaluate the performance of the structures in H_2 storage.

Keywords: Zeolites; Machine learning; Materials science

1. Introduction

In materials science, zeolites play a crucial role due to their unique properties, such as high adsorption capacity and selectivity in separation processes. The analysis and characterization of these crystalline structures have traditionally been performed using experimental methods, which has proven to be a slow and costly process with the need to explore new technologies, such as materials for storage of new fuels like hydrogen, which is becoming increasingly sought due to the need for new promising energy sources. Thus, with the advancement of data science and artificial intelligence, new approaches are being explored to optimize and accelerate the study of these substances.

One emerging method is the use of energy histograms as input parameters for machine learning algorithms [1]. Energy histograms, which represent the distribution of energies associated with different states or interactions within a structure, provide detailed information about the energy configuration of the system. This approach allows the characteristics of zeolites to be systematically quantified and transformed into data that can be processed by algorithms.

In this context, energy histograms stand out as powerful tools for capturing patterns and relationships that may be difficult to discern using traditional methods. Integrating these energy representations into machine learning algorithms can facilitate the classification, prediction, and even discovery of new zeolites with desired properties. The combination of these innovative methods promises not only to accelerate the research process but also to provide a deep perception into the structure and behavior of zeolites, significantly expanding the possibilities for application in various fields of materials science and engineering.

2. Methodology

2.1. Energy histograms using IZA Database

The energy histograms, used as inputs (descriptors) for machine learning calculations, were generated by the energy interactions from a unitary hydrogen molecule and zeolite structures, using a python automatic routine. The calculations systems were formed by a hydrogen molecular model, a Lennard-Jones center from Michels-Degraaff-Tenseldam [2], zeolites, obtained from *Database of Zeolite Structures* [3], and a set of forcefield parameters applied, Universal Forcefield (UFF) [4] with energetics parameters refined according to Gomes *et al*. [5]. The zeolite structures, obtained from the database, were replicated in the directions A, B and C until the dimension of every direction was bigger or equal to 30 Å, size superior to double the utilized cutoff (12,8 Å). The periodic contour condition was applied on the three directions to simulate the structure continuity.

We used a methodology to determinate histograms based on the works of Bucior et al. [1] e Shi et al. [6] for MOFs. Each zeolite was researched by one hydrogen molecule inserted in 1e5 random positions, tests with 1e6 random insertions and net positioning with 1,0 Å of distancing did not generate meaningful improvements on the histograms. The energy interaction from zeolite-hydrogen was calculated considering only the LJ interactions, using the mix ruling of Lorentz-Berthelot defining the cross intermolecular interactions, employing the Lennard-Jones 12-6 equation (1) being ϵ the cross-well depth, σ the crossed Van der Waals radius and r the interatomic distance of the atoms:

$$E_p\left(LJ_{12-6}\right) = \epsilon * \left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right) \qquad (1)$$

The interaction energies obtained were sorted according to their values and the quantity of energy points from each level was normalized by their frequency, obtaining histograms characterizing the energy distribution of zeolites. The first energetic level is formed with all positive potentials and the other levels formed with negative potentials with intervals of 1,0 kJ/mol (the second level, for example, covers the potentials with energetic values between 0,0 and -1,0 kJ/mol, and successively).

2.2. Machine Learning

ML models have been constantly being used to predict physical properties based on the algorithm feed. In this work we proposed the use of energy histograms as a descriptor, according to the energy potentials range [KJ/mol], to predict isotherms of adsorption of H_2 in zeolites.

Table 1. Used descriptors

Descriptor	Description
E [KJ/mol]	10 Energy potential ranges
P [KPa]	Pressure

Every descriptor passed by a normalization step Min-Max [7], to maintain the inputs in the same scale (0.1 to 0.9), according to the equation (2).

$$\alpha_n = 0.8\left(\frac{\alpha-\alpha_{min}}{\alpha_{max}-\alpha_{min}}\right) + 0.1 \qquad (2)$$

Which α_n is a normalized variable, α is a variable to be normalized and the subscripts min and max are respectively the minimum and maximum values of the descriptors in question.

Four distinct ML models associated with the libraries Scikit-learn, XGBoost e Tensorflow were evaluated by the programming language Python 3. Two of the models were based in mathematical functions: Polynomial Regression (PR) and Artificial Neural Network (ANN); and the other two were based in decision trees: Random Forest (RF) and Xtreme Gradient Boost (XGB).

3. Results and discussion

To generate the zeolite adsorption isotherms to compose the ML dataset, we previously made the forcefield refine to be used on the simulations. Validation was made in basis of the accord of simulated results and the experimental hydrogen isotherms on 298 K and well-known zeolites, just as NaX and an LSX-Li, both zeolite structures of FAU, an all-silica MFI and an LTA-Na.

The experimental isotherms reproducibility was observed when the energy parameters of Si and Al corresponded to 50% of the integral value of UFF, displayed by the Figure 1 example. This reduction can be justified by the lower interaction of adsorbate and bigger pore zeolites, have in mind the bigger the diameter of the larger sphere that can diffuse along a structure the lower is the interaction between adsorbate and adsorbent (as example, the diameter of the bigger sphere that can diffuse along LTA-Na is 4.2 Å, while the FAU zeolite this parameter is 7.35 Å.

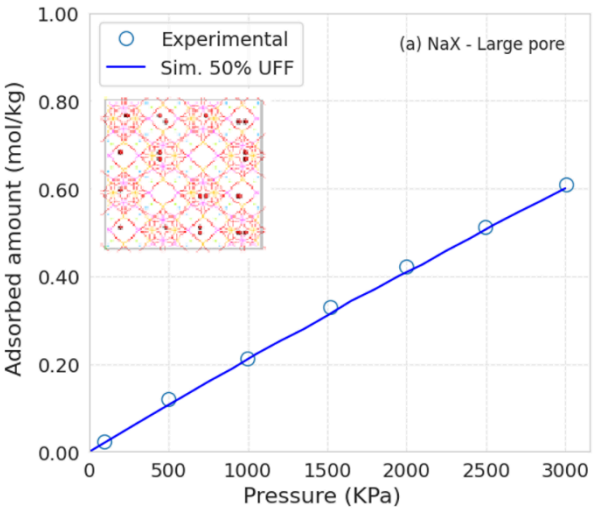

Fig. 1. Forcefield validation

The graphic results from the python routine consist of an energy frequency histogram and a 2D energy colorimetric map with Z lengths as layers, just as shown in Figures 2 and 3.

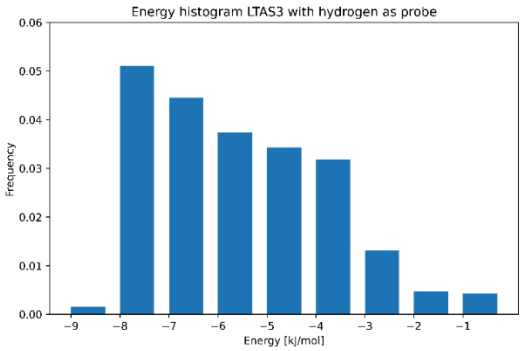

Fig. 2. Energy frequency histogram

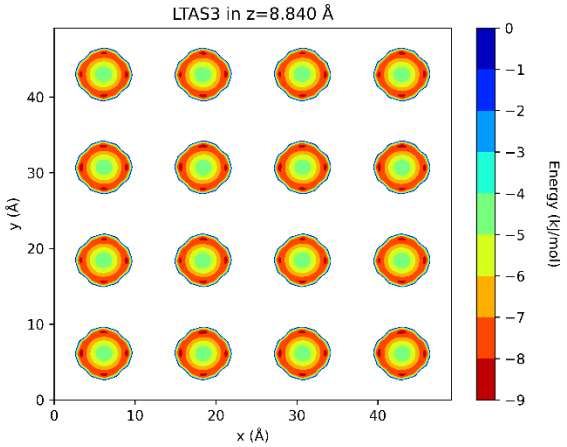

Fig. 3. 2D energy colorimetric map

To train these four models with energy histograms, we checked the accuracy by the accumulated frequency percentage of absolute relative error of the test data (Figure 4). The three models with best performance observed: RF, XGB and ANN, having emphasis in the XGBoost model in accord to the decision tree, with around 90% of the predicted data displaying relative error absolute inferior to 15%.

A diagram of equality was plotted between the simulated data and the predicts by the XGBoost algorithm, was possible to verify low dispersion points around the bissetrix, with determination parameter (R^2) equal to 0.985 on the test group (Figure 5).

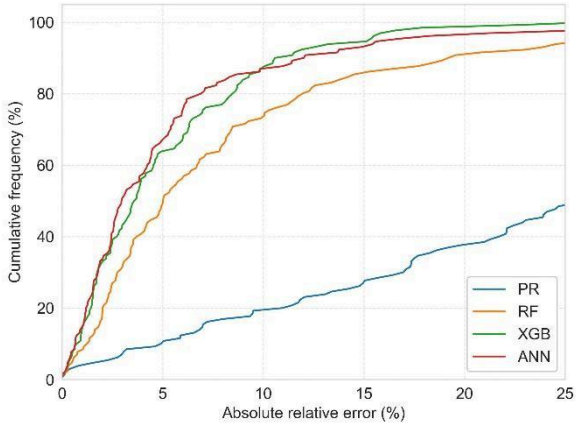

Fig 4. Accumulated frequency – test data.

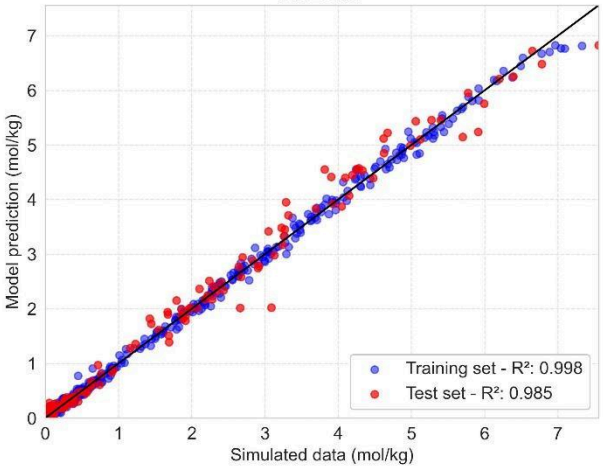

Fig 5. Prediction results for the XGB model.

4. Conclusion

On the analysis of the energy histograms made for 42 zeolites structures of the IZA database was noticed that the structures of LTAS3 and NaX showed a higher frequency on their energy distribution.

The models ANN and XGB were displayed satisfactory statically to predict the adsorption isotherms on zeolites. Therefore, the trained model has the capacity to evaluate the performance of the structures on the IZA database to store H_2, significantly reducing the costs of time and resources on the choice of the better material.

Acknowledgements

The authors would like to thank FUNCAP, CAPES, and CNPq for the financial support and the use of computer clusters at National Laboratory of Scientific Computing (LNCC/MCTI, Brazil).

References

[1] B. J. Bucior *et al.*, "Energy-based descriptors to rapidly predict hydrogen storage in metal–organic frameworks," *Mol. Syst. Des. Eng.*, vol. 4, no. 1, pp. 162–174, 2019, doi: 10.1039/C8ME00050F.

[2] A. Michels, W. de Graaff, and C. A. Ten Seldam, "Virial coefficients of hydrogen and deuterium at temperatures between −175°C and +150°C. Conclusions from the second virial coefficient with regards to the intermolecular potential," Physica, vol. 26, no. 6, pp. 393–408, Jun. 1960, doi: 10.1016/0031-8914(60)90029-X.

[3] C. Baerlocher, L. McCusker, and D. Olson, Atlas of Zeolite Framework Types. Elsevier, 2007. doi: 10.1016/B978-0-444-53064-6.X5186-X.

[4] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations," J. Am. Chem. Soc., vol. 114, no. 25, pp. 10024–10035, Dec. 1992, doi: 10.1021/ja00051a040.

[5] V. A. M. Gomes, J. A. Coelho, H. R. Peixoto, and S. M. P. Lucena, "Easily tunable parameterization of a force field for gas adsorption on FAU zeolites," *Adsorption*, vol. 21, no. 1–2, pp. 25–35, Feb. 2015, doi: 10.1007/s10450-014-9647-3.

[6] K. Shi *et al.*, "Two-Dimensional Energy Histograms as Features for Machine Learning to Predict Adsorption in Diverse Nanoporous Materials," *J. Chem. Theory Comput.*, vol. 19, no. 14, pp. 4568–4583, Jul. 2023, doi: 10.1021/acs.jctc.2c00798.

[7] A. R. Khataee and M. B. Kasiri, "Artificial neural networks modeling of contaminated water treatment processes by homogeneous and heterogeneous nanocatalysis," *J. Mol. Catal. A Chem.*, vol. 331, no. 1–2, pp. 86–100, Oct. 2010, doi: 10.1016/j.molcata.2010.07.016.