Supplemental Information for Graph Neural Network-Based Multi-Objective Bayesian Optimization for Enhanced Screening of Metal–Organic Frameworks with Optimal Separation Performance

Lane E. Schultz[1,*], Nickolas Gantzler[2], N. Scott Bobbitt[2], Dorina F. Sava Gallis[3], and Remi Dingreville[1]

[1]Center for Integrated Nanotechnologies (CINT), Sandia National Laboratories, Albuquerque, NM 87185, USA.
[2]Material, Physical, and Chemical Sciences Center, Sandia National Laboratories, Albuquerque, NM 87185, USA.
[3]Nanoscale Sciences Department, Sandia National Laboratories, Albuquerque, NM 87185, USA.
[*]Corresponding author: leschul@sandia.gov

January 14, 2026

1 Kinetic diameters

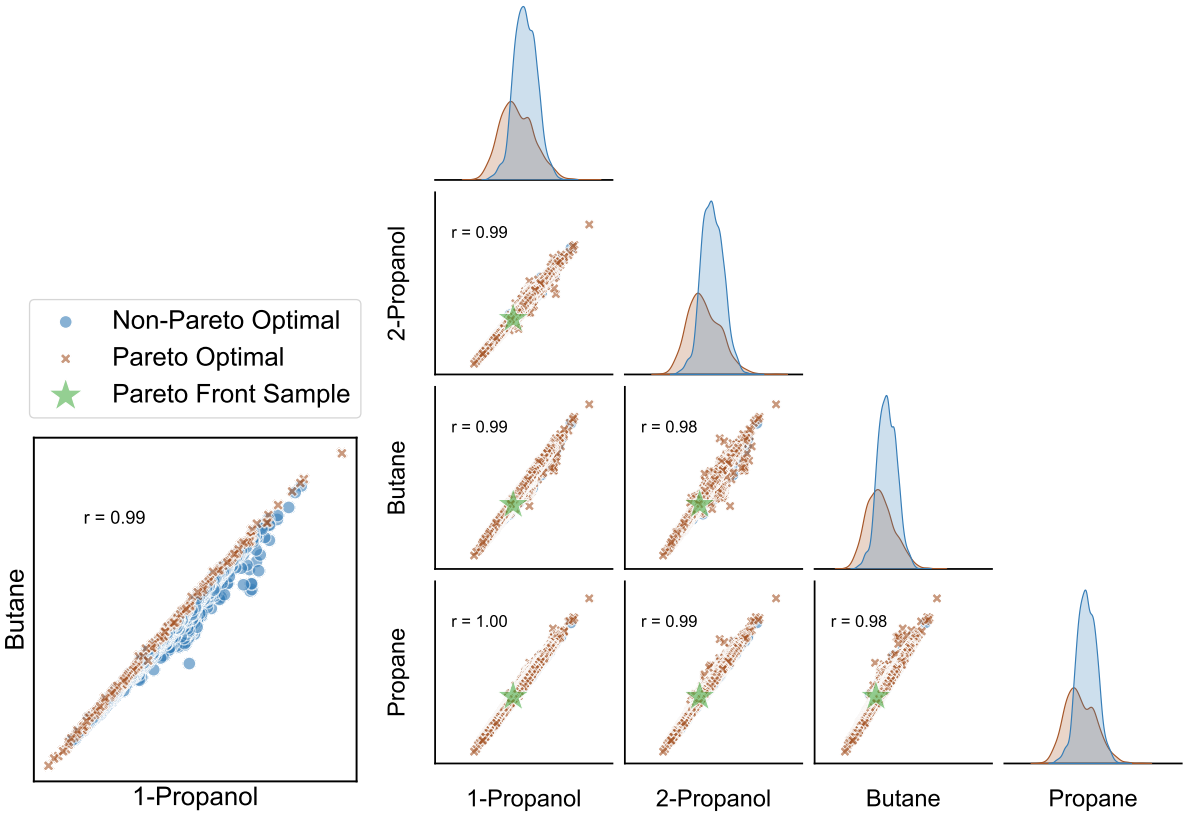
The kinetic diameters of volatile organic compounds, which provide a lower bound estimate for the pore limiting diameter of the studied metal-oxide frameworks, are tabulated in Table 1.

Table 1: Kinetic diameters used as cutoffs for the pore limmiting diameter

Name	Diameter (Å)	Reference	DOI URL
1-propanol	4.7	1	https://doi.org/10.1039/C5CC00113G
2-propanol	4.7	1	https://doi.org/10.1039/C5CC00113G
2-cis-butene	4.94	2	https://doi.org/10.1002/cssc.201700657
2-trans-butene	4.31	2	https://doi.org/10.1002/cssc.201700657
benzene	5.9	3	https://doi.org/10.3390/ma15176100
toluene	5.9	4	https://doi.org/10.1016/j.memsci.2018.05.016
butane	4.687	2	https://doi.org/10.1002/cssc.201700657
propane	4.3	5	https://doi.org/10.1016/j.micromeso.2020.110099

2 Select correlation plots for Henry coefficients

A select number of Henry coefficients for VOCs are shown in Fig. 1. These correlation plots illustrate the difficulty in optimizing competing objectives (i.e., minimizing the Henry coefficient for 1-propanol while maximizing that of butane). Since the visualization is limited to two-dimensional slices, some Pareto-optimal points overlap visually with suboptimal ones.



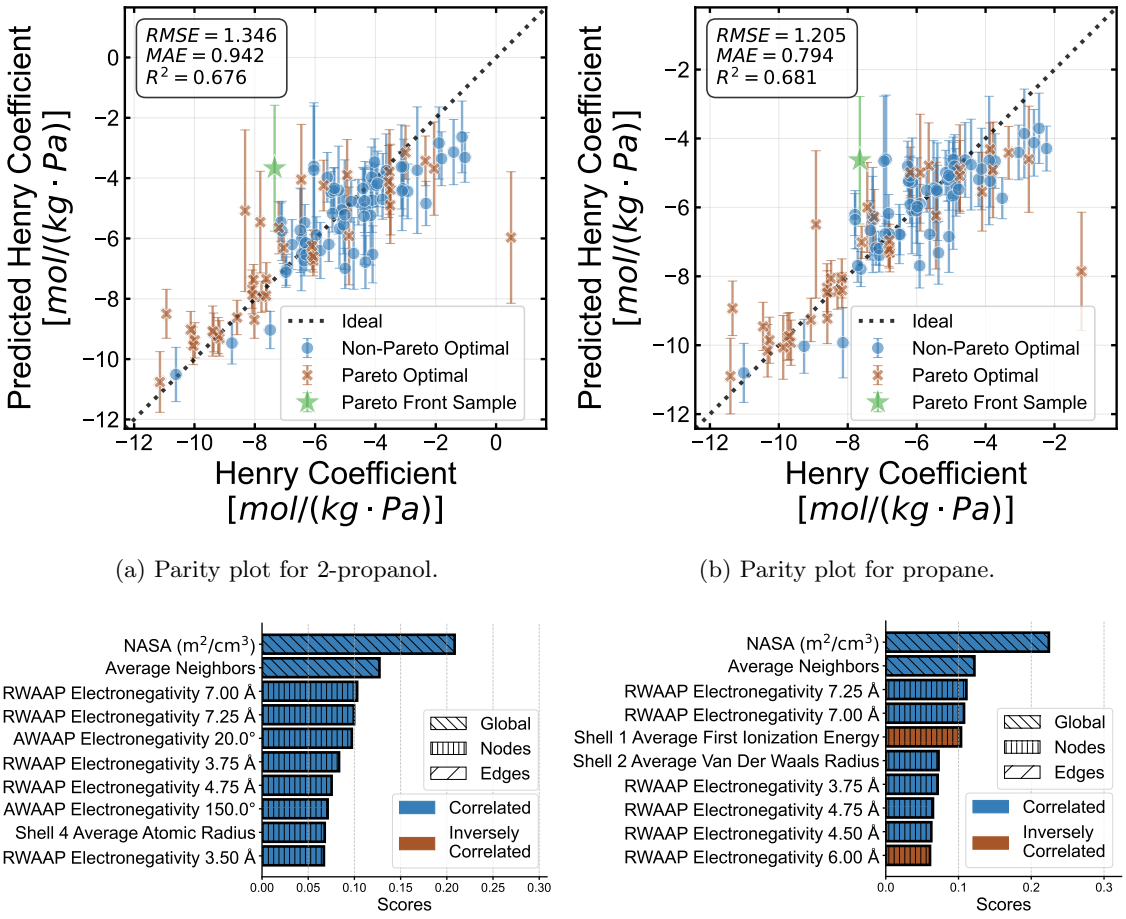
(a) Two objectives.

(b) Four objectives.

Figure 1: Correlations among optimization objectives. Panel (a) shows a two objective Pareto front where the Henry coefficients for butane are maximized while 1-propanol are minimized The optimal points are to the top left of the others. Panel (b) shows the Pareto front for optimizing four Henry coefficients: butane (maximized), propane (maximized), 1-propanol (minimized), and 2-propanol (minimized). Each subplot displays one objective on the horizontal axis, its correlation with a second objective on the vertical axis, and the Pearson correlation factor (r). Blue circles denote non-Pareto-optimal solutions, brown crosses indicate Pareto-optimal solutions, and the green star highlights a sample used for feature sensitivity analysis in the main text. The green star point lies on the Pareto front for optimizing four objectives. Distributions show data densities.

3 Remaining parity plots and ablation study for four objective optimization

Fig. 2 shows the remaining parity and ablation plots not included in the main manuscript.



(a) Parity plot for 2-propanol.

(b) Parity plot for propane.

(c) Ablation for 2-propanol.

(d) Ablation for propane.

Figure 2: Parity plots and feature ablation analysis. Panels (a)–(b) present parity plots illustrating the performance of a single surrogate model in predicting each objective. The units are in a natural logarithm. The single green point represents the Pareto front sample used to explain feature significance. The blue circles and brown crosses are the non-Pareto front and Pareto front points, respectively. Panels (c)–(d) display the top 10 features from a feature ablation study for the green star point. In these plots, blue bars indicate that increasing a feature value during ablation leads to an increase in the model prediction (i.e., correlated), while brown bars indicate that increasing the feature value results in a decrease in the model prediction (i.e., inversely correlated). Bar patterns correspond to the type of graph feature (e.g., global, node, or edge). Scores were normalized such that the sum of all is one.

4 Gaussian process in Bayesian Optimization

We wrote a workflow using Gaussian process (GP) as a surrogate for Bayesian optimization (BO). The code was validated against the work in Deshwal et al.[6] for covalent organic frameworks (COFs). They averaged their results with 100 independent BO runs (Fig. 5 (a))[6] while we only used 10. Nevertheless, all runs from our work and the referenced publication reach the highest objective value after around 110 evaluated COFs. The main difference between results was the shaded areas due to the lower number of sampling conducted in our work. Please refer to Fig. 5 (a) from Deshwal et al.[6] for their results. Our results are shown in Fig. 3.

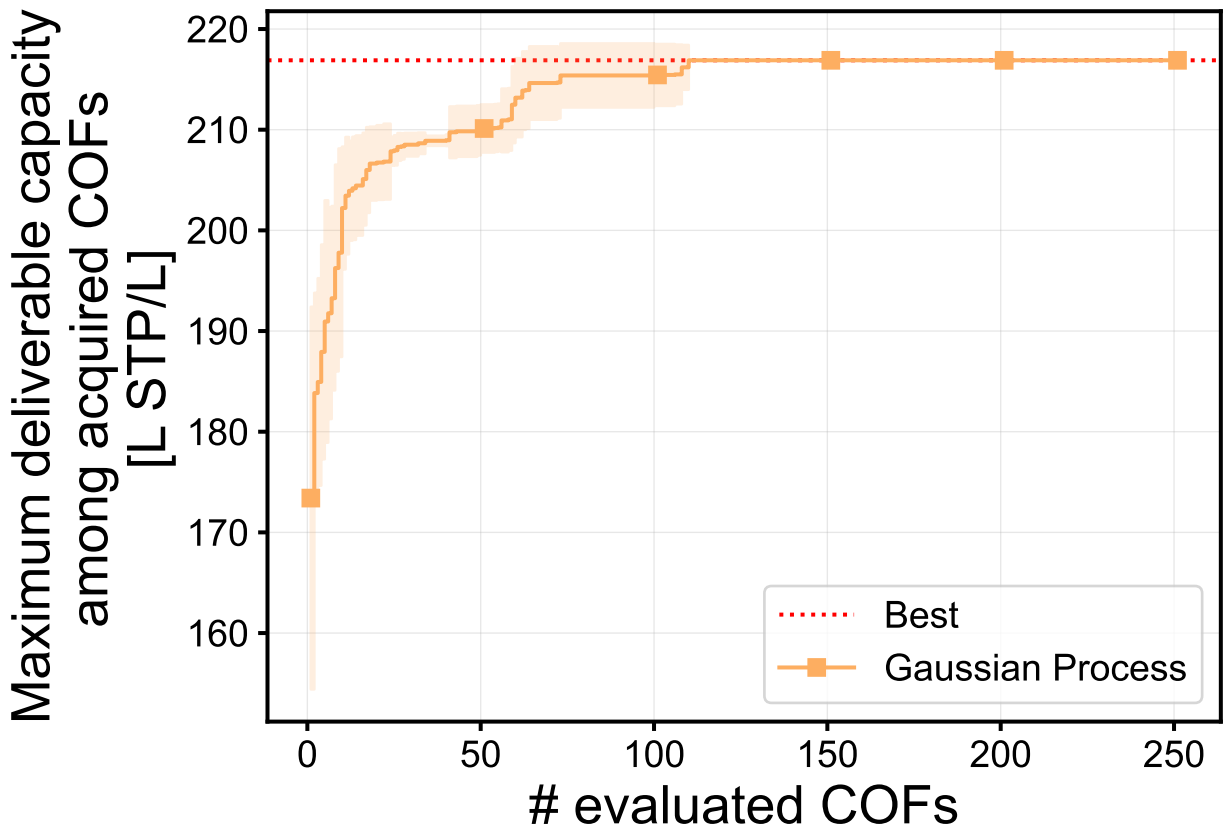


Figure 3: A Bayesian optimization run for selecting covalent organic frameworks with higher deliverable capacities using Gaussian possess as a surrogate model. The dotted line is the highest deliverable capacity in the dataset. The line with the square markers are the mean from selections with Gaussian process surrogates and the shaded area is the standard deviation.

References

[1] R. K. Motkuri, P. K. Thallapally, H. V. R. Annapureddy, L. X. Dang, R. Krishna, S. K. Nune, C. A. Fernandez, J. Liu and B. P. McGrail, *Chemical Communications*, 2015, **51**, 8421–8424.

[2] M. Gehre, Z. Guo, G. Rothenberg and S. Tanase, *ChemSusChem*, 2017, **10**, 3947–3963.

[3] W. Chen, H. Zhao, Y. Xue and X. Chang, *Materials*, 2022, **15**, 6100.

[4] N. O. Chisholm, H. H. Funke, R. D. Noble and J. L. Falconer, *Journal of Membrane Science*, 2018, **560**, 108–114.

[5] D. Saha, B. Toof, R. Krishna, G. Orkoulas, P. Gismondi, R. Thorpe and M. L. Comroe, *Microporous and Mesoporous Materials*, 2020, **299**, 110099.

[6] A. Deshwal, C. M. Simon and J. R. Doppa, *Molecular Systems Design & Engineering*, 2021, **6**, 1066–1086.