Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/20/eaaz4301/DC1

Supplementary Materials for

Designing exceptional gas-separation polymer membranes using machine learning

J. Wesley Barnett, Connor R. Bilchak, Yiwen Wang, Brian C. Benicewicz,
Laura A. Murdock, Tristan Bereau, Sanat K. Kumar*

*Corresponding author. Email: sk2794@columbia.edu

Published 15 May 2020, *Sci. Adv.* **6**, eaaz4301 (2020)
DOI: 10.1126/sciadv.aaz4301

This PDF file includes:

Table S1: Model performance using the train set/test set approach described in the main text. The mean square error (MSE) in the test set data is included.

Gas	Full size	Train size	Train R^2	Test size	Test R^2	Test MSE
N_2	686	514	0.986	172	0.847	0.218
O_2	698	523	0.985	175	0.903	0.134
H_2	433	324	0.985	109	0.827	0.163
He	376	282	0.980	94	0.799	0.150
CH_4	561	420	0.990	141	0.904	0.219
CO_2	629	471	0.986	158	0.875	0.187

Table S2: Likelihood ($\mathcal{L}$) and hyperparameters for models trained using the split training set and the full data set. σ_f is the variance in the signal, ℓ is the length scale of the radial basis function, and $\sigma_n{}^2$ Is the standard deviation of the noise.

Gas	Training data set				Full data set			
	$\mathcal{L}$	σ_f	ℓ	σ_n^2	$\mathcal{L}$	σ_f	ℓ	σ_n^2
CH_4	-372.9	6.75	201.0	0.0594	-441.0	2.98	149.0	0.0508
CO_2	-356.1	2.66	95.3	0.0514	-422.7	2.78	99.7	0.0488
He	-141.2	2.06	114.0	0.0350	-167.3	2.01	106.0	0.0323
H_2	-171.3	1.98	100.0	0.0365	-216.1	2.00	101.0	0.0442
N_2	-449.7	2.58	83.1	0.0664	-521.3	2.76	90.5	0.0585
O_2	-407.4	2.81	101.0	0.0559	-460.6	2.91	105.0	0.0473

Table S3: Breakdown of polymer classes for all 11,325 polymers predicted and for the 100 polymers the above upper bound in CO_2/CH_4 Robeson plot.

Group	% all predicted	% above upper bound
Polyacrylics	6.78	1.00
Polyamides/thioamides	21.50	20.00
Polyanhydrides/thioanhydrides	1.17	0.00
Polycarbonates/thiocarbonates	2.58	0.00
Polydienes	1.00	0.00
Polyesters/thioesters	18.53	8.00
Polyhalo-olefins	0.42	0.00
Polyimides/thioimides	17.65	35.00
Polyimines	17.54	9.00
Polyketones/thioketones	7.97	1.00
Polyolefins	0.86	0.00
Polyoxides/ethers/acetals	30.78	21.00
Polyphenylenes	2.04	0.00
Polyphosphazenes	0.61	0.00
Polysiloxanes/silanes	3.58	0.00
Polystyrenes	2.43	0.00
Polysulfides	6.99	53.00
Polysulfones/sulfoxides/sufonates/sulfoamides	5.30	18.00
Polyureas/thioureas	1.76	0.00
Polyurethanes/thiourethanes	3.71	0.00
Polyvinyls	13.42	1.00
Other polymers	0.67	0.00

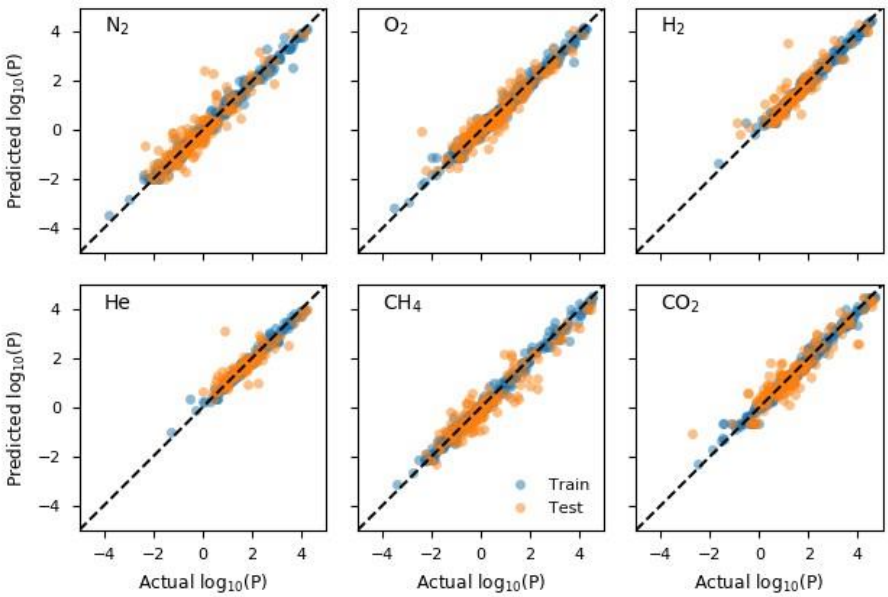

Figure S1: Plots of the actual base-10 logarithm of permeabilities versus the predicted values from the trained models fit on the training sets. P is in units of Barrer.

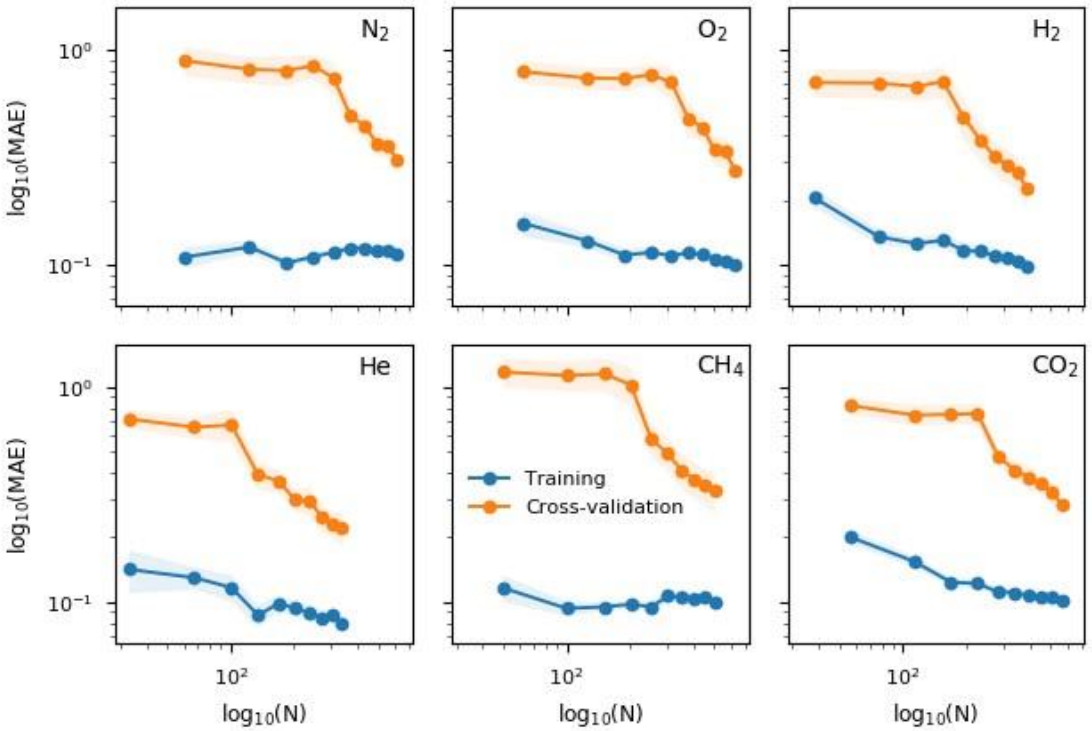

Figure S2: Learning curves for each model using the full data set. Training and cross-validation scores are logarithm (base 10) of the mean absolute error as a function of the logarithm of the number of samples. The shaded regions indicate the standard deviation from the 10-fold cross-validation.