

## Prediction of CO<sub>2</sub> absorption by physical solvents using a chemoinformatics-based machine learning model

Hao Li <sup>a</sup>, Dan Yan <sup>b</sup>, Zhien Zhang <sup>c\*</sup>, Eric Lichtfouse <sup>d</sup>

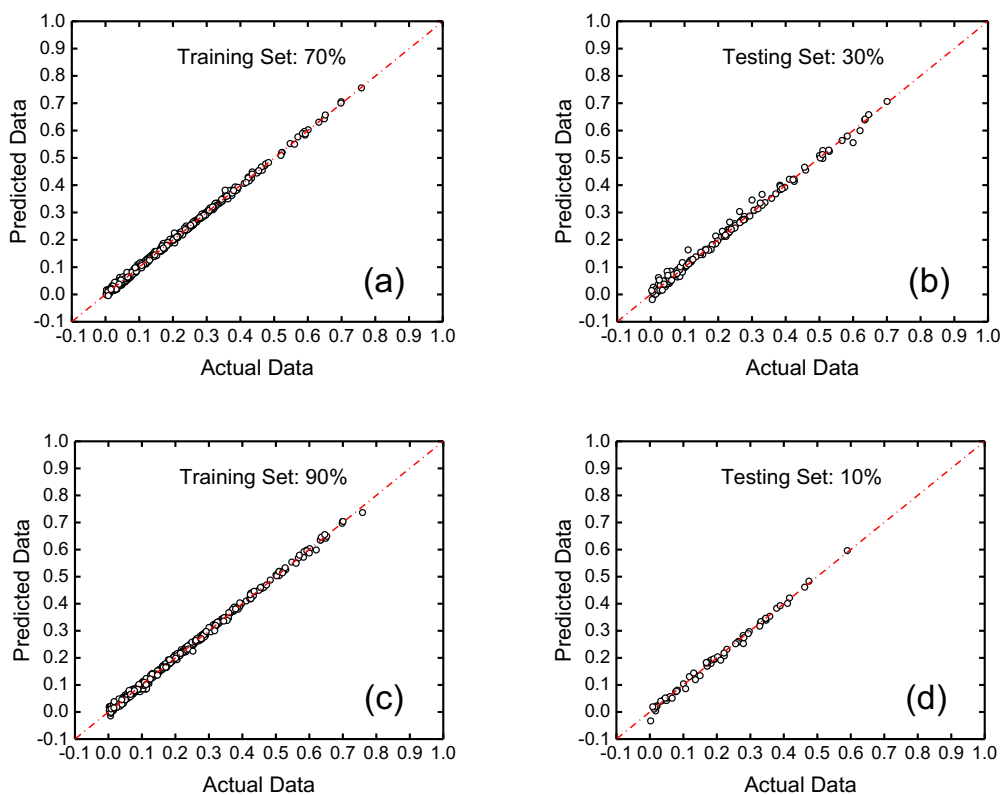
<sup>a</sup> Department of Chemistry and the Institute for Computational and Engineering Sciences, The University of Texas at Austin, 105 E. 24th Street, Stop A5300, Austin, TX 78712, USA;

<sup>b</sup> Shenzhen Environmental Science and Technology Engineering Laboratory, Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, China;

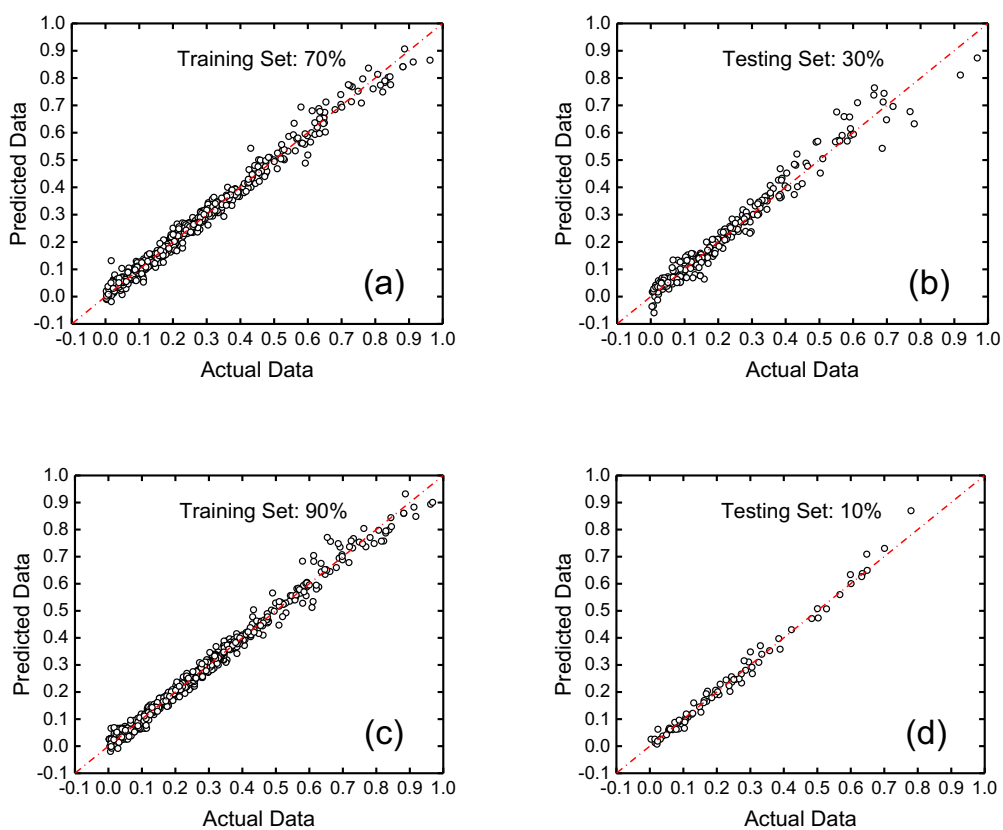
<sup>c</sup> William G. Lowrie Department of Chemical and Biomolecular Engineering, The Ohio State University, Columbus, Ohio 43210, USA;

<sup>d</sup> CEREGE, Aix-Marseille Univ, Coll de France, CNRS, INRA, IRD, Aix-en-Provence, France.

\* Email: [zhienzhang@hotmail.com](mailto:zhienzhang@hotmail.com); [zhang.4528@osu.edu](mailto:zhang.4528@osu.edu) (Z. Zhang)



**Figure S1.** Selected typical training and testing results of predicted values versus actual values trained and tested from 624 data groups extracted from [1]. (a,b) 70% training + 30% testing, and (c,d) 90% training + 10% testing. The diagonal dashed line represents the line of  $y=x$ . The data point which is close to the diagonal dashed line indicates a good prediction.



**Figure S2.** Selected typical training and testing results of predicted values versus actual values trained and tested from 823 data. (a,b) 70% training + 30% testing, and (c,d) 90% training + 10% testing. The diagonal dashed line represents the line of  $y=x$ . The data point which is close to the diagonal dashed line indicates a good prediction.

### Reference

I. Gui, X.; Tang, Z.; Fei, W. Solubility of  $\text{CO}_2$  in alcohols, glycols, ethers, and ketones at high pressures from (288.15 to 318.15) K. *J. Chem. Eng. Data* **2011**, *56*, 2420–2429, DOI:10.1021/je101344v.