



**HAL**  
open science

# Improved reverse engineering based on phase equilibria prediction for CO<sub>2</sub> capture by integrating deep learning to COSMO-RS

Edoardo Parascandolo, Vincent Gerbaud, Sophie Thiebaud-Roux, Ivonne Rodriguez-Donis

## ► To cite this version:

Edoardo Parascandolo, Vincent Gerbaud, Sophie Thiebaud-Roux, Ivonne Rodriguez-Donis. Improved reverse engineering based on phase equilibria prediction for CO<sub>2</sub> capture by integrating deep learning to COSMO-RS. 15th European Congress of Chemical Engineering (ECCE), Sep 2025, Lisbon (Portugal), Portugal. ⟨hal-05135961⟩

**HAL Id: hal-05135961**

**<https://hal.science/hal-05135961v1>**

Submitted on 5 Nov 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# Improved reverse engineering based on phase equilibria prediction for CO<sub>2</sub> capture by integrating deep learning to COSMO-RS

*Edoardo Parascandolo<sup>\*</sup>, Vincent Gerbaud<sup>\*\*</sup>, Sophie Thiebaud-Roux<sup>\*</sup>, Ivonne Rodriguez-Donis<sup>\*</sup>*

*<sup>\*</sup>Laboratoire de Chimie Agro-industrielle, Toulouse-INP, Toulouse/FR*

*<sup>\*\*</sup>Laboratoire de Génie Chimique, Toulouse-INP, Toulouse/FR*

Solvent selection is a critical step in designing efficient gas separation processes. In contrast with traditional trial-and-error methods, reverse engineering allows for a targeted exploration of a wider list of candidates, reducing reliance on empirical methods and accelerating the development of effective solutions. In a three sequential phases approach, namely “modelisation”, “evaluation/design” and “selection”, a big list of existing solvents can be analyzed through benchmarking to identify relevant properties and applications. By defining a list of target properties, the relationship between molecular structure and performance is exploited through the use computational modeling to identify new solvent candidates. Nevertheless, the need for fully-predictive estimations in both the “modelisation” and “evaluation/design” phase can substantially impact the quality of the results, potentially turning into misleading conclusion. Our study deals with the computation of phase diagrams for binary mixtures of CO<sub>2</sub> with candidate solvents for physical absorption process using the quantum chemical-based thermodynamic model Conductor like Screening MOdel for Real Solvents (COSMO-RS) [1] to obtain vapor liquid equilibrium curves as first measure of solubility. This demand cannot be addressed by the sole use of COSMO-RS due to its assumption of ideal gas phase and dependencies of the activity coefficient model on quantum chemistry descriptors, that can lead to high deviations when handling complex molecular interactions or non-ideal systems [2]. To handle these limitations, we propose the integration of a message passing graph-convolutional neural network [3], [4] to the classic methodology. Due to a consistent lack in experimental data for sufficiently training a robust model, equilibrium curves estimated at three temperatures (0 °C, 25 °C, 50 °C) in COSMO-RS for binary mixtures of CO<sub>2</sub> with 1241 liquid molecules (at room temperature) without reactive groups were revised and collected in a first database to be used for pre-training the neural network.

Finally, experimental data encompassing 43 different physical solvent were retrieved via the Nist ThermoData Engine SRD#103b v10.4.3 [5] and used in a milder fine-tuning phase, where most of the parameters of both the message passing and the feed forward network (MPNN, FNN) were constrained to avoid catastrophic inference.

To include operating conditions to the model, weighted values of liquid molar fractions and equilibrium temperatures are added as further inputs for a scaling the images of each molecular embedding produced after the MPNN. Both 3x3 repeated cross validation and early stopping are implemented to avoid overfitting, while successive Bayesian algorithms have been employed for optimizing the hyperparameters. The results obtained by averaging the values of a 9-model-ensemble showed a significant reduction in prediction errors: in comparison with COSMO-RS estimations, the mean absolute error (MAE) decreased from 7.07 bar to 3.34 bar while the Pearson correlation coefficient ( $R^2$ ) in the parity plot increased from 0.88 to 0.955. In conclusion, this refined approach, combining the strengths of machine learning and physics-based models, provides a scalable and accurate tool for phase equilibria calculations, that can be furtherly included in an enhanced reverse engineering approach for finding novel solvents for CO<sub>2</sub> capture by physical absorption process.

## Bibliography

- [1] C. C. Pye, T. Ziegler, E. van Lenthe, and J. N. Louwen, "An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package — Part II. COSMO for real solvents," *Can J Chem*, vol. 87, no. 7, pp. 790–797, Feb. 2009, doi: 10.1139/V09-008.
- [2] A. Klamt, "Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena," *J Phys Chem*, vol. 99, no. 7, pp. 2224–2235, Feb. 1995, doi: 10.1021/j100007a062.
- [3] E. Heid *et al.*, "Chemprop: A Machine Learning Package for Chemical Property Prediction," *J Chem Inf Model*, vol. 64, no. 1, pp. 9–17, Jan. 2024, doi: 10.1021/acs.jcim.3c01250.
- [4] K. Yang *et al.*, "Analyzing Learned Molecular Representations for Property Prediction," *J Chem Inf Model*, vol. 59, no. 8, pp. 3370–3388, Aug. 2019, doi: 10.1021/acs.jcim.9b00237.
- [5] NIST, "NIST Standard Reference Database 103b: ThermoData Engine," 2022. [Online]. Available: <https://www.nist.gov/srd/nist-standard-reference-database-103b>