Computers and Chemical Engineering 211 (2026) 109678

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/cace

Short communication

Clapeyron neural networks for single-species vapor–liquid equilibria

Jan Pavšek [a], Alexander Mitsos [b,a,c], Elvis J. Sim [a], Jan G. Rittig [a,*]

[a] Process Systems Engineering (AVT.SVT), RWTH Aachen University, Forckenbeckstraße 51, 52074 Aachen, Germany
[b] JARA-CSD, Templergraben 55, 52056 Aachen, Germany
[c] Institute of Climate and Energy Systems ICE-1: Energy Systems Engineering, Forschungszentrum Jülich GmbH, Wilhelm-Johnen-Straße, 52425 Jülich, Germany

ARTICLE INFO

Dataset link: NIST Thermodata engine

Keywords:
Property prediction
Thermodynamics
Machine learning
Graph neural network
vapor–liquid equilibrium

ABSTRACT

Machine learning (ML) approaches have shown promising results for predicting molecular properties relevant for chemical process design. However, they are often limited by scarce experimental property data and lack thermodynamic consistency. As such, thermodynamics-informed ML, i.e., incorporating thermodynamic relations into the loss function as regularization term for training, has been proposed. We herein transfer the concept of thermodynamics-informed graph neural networks (GNNs) from the Gibbs–Duhem to the Clapeyron equation, predicting several pure component properties in a multi-task manner, namely: vapor pressure, liquid molar volume, vapor molar volume and enthalpy of vaporization. We find improved prediction accuracy of the Clapeyron-GNN compared to the single-task learning setting, and improved approximation of the Clapeyron equation compared to the purely data-driven multi-task learning setting. In fact, we observe the largest improvement in prediction accuracy for the properties with the lowest availability of data, making our model promising for practical application in data scarce scenarios of chemical engineering practice.

1. Introduction

Process design requires information on multiple thermodynamic properties for the species included in the process, such as density, enthalpy and vapor pressure. Here, molecular ML has shown to provide accurate predictions for a variety of such properties, including both pure component properties, e.g., density (Winter et al., 2025) and phase transition enthalpy (Leenhouts et al., 2025), and mixtures, especially for activity coefficients (Qin et al., 2023; Rittig and Mitsos, 2024; Specht et al., 2024; Medina and Sundmacher, 2026; Wahyudi et al., 2026).

Recently, research has particularly focused on advancing molecular ML by incorporating thermodynamic relations into the model architecture and training, in the form of hybrid and physics-informed ML, see overview in Rittig et al. (2026) and Wu et al. (2023). Thermodynamics-informed approaches are particularly promising, as they do not rely on semi-empirical thermodynamic models, which introduce corresponding modeling limitations and assumptions; rather, they use differential equations to thermodynamic potentials. These thermodynamics-enriched ML approaches reduce the data required for training and enhance — in some cases, even guarantee — thermodynamic consistency of the predictions (Rosenberger et al., 2022; Rittig and Mitsos, 2024; Specht et al., 2024), making them highly promising for chemical process design applications. However, most thermodynamics-informed ML models focus on predicting a single property of interest.

We transfer the concept of thermodynamics-informed ML to single-species vapor–liquid equilibria prediction and extend it with multi-task training. That is, we train a GNN to predict vapor and liquid molar volume, vapor pressure and enthalpy of vaporization of a single-species system at vapor–liquid equilibrium. We jointly train on all four properties in a multi-task learning setting. The motivation for this lies in the correlation between the four properties, which is expected to improve predictive performance by means of multi-task learning (Sosnin et al., 2019; Liu et al., 2021). We use the Clapeyron equation as physics regularization (Raissi et al., 2019) in the training loss function to enhance physical consistency of the network predictions, i.e., following a thermodynamics-informed approach. Notably in preliminary studies, we also tested a thermodynamics-consistent approach by directly embedding the Clapeyron equation into the GNN architecture. This, however, resulted in significantly lower prediction accuracy, which we attribute to the imbalanced dataset of the four independently measured properties making model training challenging, also cf. Alam et al. (2026). The Clapeyron-informed approach is more flexible here — albeit at the expense of guaranteed consistency — as the Clapeyron equation acts as a soft constraint in the training loss and can be adapted using a weighting factor. We refer to our architecture as Clapeyron-GNN.

A comparable approach has recently been published by Park et al. (2025), who directly predict single-species vapor–liquid equilibria with

* Corresponding author.
E-mail address: jan.rittig@rwth-aachen.de (J.G. Rittig).

https://doi.org/10.1016/j.compchemeng.2026.109678
Received 24 February 2026; Received in revised form 14 April 2026; Accepted 18 April 2026
Available online 21 April 2026

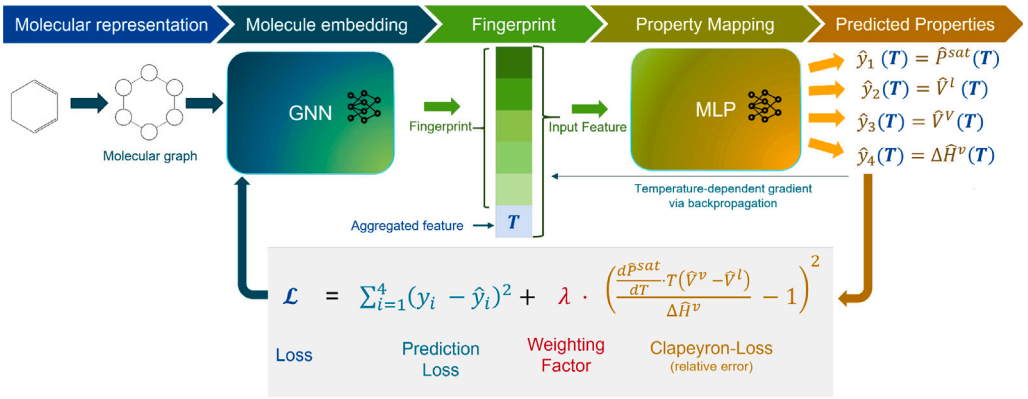


Fig. 1. Schematic illustration of Clapeyron-GNN.

a GNN. However, they not only provide the graph structure of the molecules as input but also additional graph level features such as the acentric factor, which might not necessarily be available when making predictions for new molecules. Additionally, Kochi et al. (2025) employ the Clausius–Clapeyron equation for physics-informed vapor pressure prediction, using *chemprop* as graph encoder (Heid et al., 2023), but only predict vapor pressure and dynamic viscosity. With Clapeyron-GNN, we train an architecture that predicts all four properties, vapor pressure, liquid and vapor molar volumes and enthalpy of vaporization, in a multi-task learning setting and relies purely on the molecular structure and temperature, and thereby enables easier prediction for new molecules. We benchmark the predictive performance of Clapeyron-GNN against a purely data-driven multi-task GNN and against four individual GNNs predicting all four properties in a single-task learning setting.

2. Clapeyron graph neural network

The Clapeyron-GNN architecture is based on our previously employed GNNs (Schweidtmann et al., 2020). That is, we map molecular graphs via graph convolutional layers and a multi-layer perceptron to molecular properties. The prediction targets are the temperature-dependent vapor and liquid molar volume, the vapor pressure and the enthalpy of vaporization (see Fig. 1). Temperature is considered as additional input by concatenation with the molecular fingerprint.

We incorporate the Clapeyron equation in the GNN model training. The Clapeyron equation is an exact equation derived from the equality of Gibbs free energy in the vapor and liquid phase; it provides a relationship between the four target properties and the temperature:

$$\frac{dp^{sat}}{dT} = \frac{\Delta H_V}{T(V^V - V^L)} \tag{1}$$

where p^{sat} is the vapor pressure, ΔH_V is the enthalpy of vaporization, V^V is the vapor molar volume and V^L is the liquid molar volume. All four properties depend on the temperature T. We reformulate the Clapeyron equation to a Clapeyron error, which measures relative deviation of the network predictions from the nearest point fulfilling the Clapeyron equation:

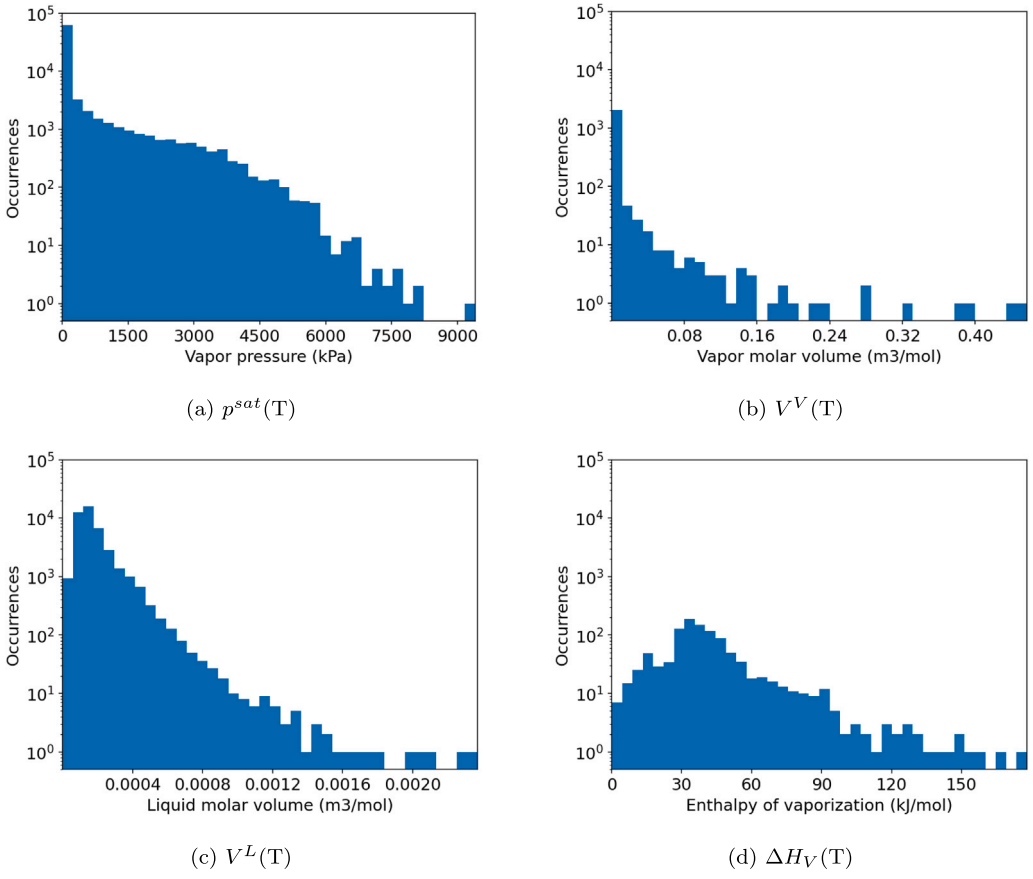
$$\mathcal{L}_{\text{Clapeyron}} = \left(\frac{\frac{d\hat{p}^{sat}}{dT} T(\hat{V}^V - \hat{V}^L)}{\Delta\hat{H}_V} - 1\right)^2. \tag{2}$$

Similarly to our previous thermodynamics-informed approach (Rittig et al., 2023), we extend the training loss by adding the Clapeyron error as additional regularization to the prediction error using a weighting factor λ, see Fig. 1. The gradient of the vapor pressure with respect to the temperature is obtained during training from backpropagation, preserving end-to-end learning. In contrast to the prediction loss, which can only be evaluated for each target property if data exists for the particular property, the Clapeyron regularization provides a loss signal for all four properties, even for temperature points for which only data on a subset of the target properties is available. This is also a major practical difference between employing a thermodynamics-informed and a thermodynamics-consistent approach. As thermodynamics-consistent models embed the thermodynamics equations directly in the output head of the model, they rely on the same number of loss signals as purely data-driven models. We performed first investigations of employing a thermodynamics-consistent approach, ie., directly embedding the exact Clapeyron equation into the output head, which resulted in non-converging training and lower prediction performance than purely data-driven models. Hence, in the present work we employ the thermodynamics-*informed* Clapeyron-GNN.

3. Case study

3.1. Dataset

We extract experimental data for the vapor and liquid molar volume, the vapor pressure and the enthalpy of vaporization from the NIST ThermoData Engine (National Institute of Standards and Technology (NIST), 2025). Notably, we manually removed 10 outliers, whose numeric values deviated by at least an order of magnitude from the rest of the dataset. Our dataset contains 879 molecules at varying temperatures, ranging from 56.75 K to 1021 K, with 102,121 data-points in total. The molecules are organic compounds, including amines, esters, alcohols, carboxylic acids, ketones, phenols, amines, nitro compounds, amides, chlorine and fluorine compounds, and range in molecular weight from 27 to 493 $\frac{g}{mol}$. The distribution of data points across different molecules and the four different properties is highly uneven (see Fig. 2). Specifically, the amount of data points varies from a single point to more than a thousand for different molecules. Considering the properties, our data sets contains 78,840 data points for the vapor pressure, 43,056 for liquid molar volume, 2206 for vapor molar volume and 1057 for enthalpy of vaporization. In fact, there is abundant training data across the temperature domain for the vapor pressure and the liquid molar volume, whereas for the vapor molar volume and the enthalpy of vaporization, most molecules have either a single or no data point at all. Therefore, learning the temperature dependency for these two properties is inherently more difficult. Not only does the number of available datapoints vary largely between properties, but also within one property, the distribution of numeric values is skewed (see Fig. 2). To account for this, all four properties are predicted on a logarithmic scale; temperature values are linearly normalized to range 0–1. All error metrics are therefore reported on the logarithmic scale.



(a) $p^{sat}(\mathrm{T})$ (b) $V^{V}(\mathrm{T})$

(c) $V^{L}(\mathrm{T})$ (d) $\Delta H_{V}(\mathrm{T})$

Fig. 2. Histograms of data distributions for four target properties and 40 bins for numeric values across all properties.

3.2. Prediction scenario & benchmark

We evaluate the prediction performance for extrapolating to new molecules. Hence, we randomly select all 80% of the molecules with corresponding data points for the training set and keep the remaining 20% of molecules and associated data points for model testing. To evaluate the contribution of the multi-task learning setting and the physics regularization, we compare Clapeyron-GNN against two base cases: (i) a purely data driven multi-task learning setting, which we refer to as MTL-GNN, and (ii) a purely data driven single-task learning setting, where we train separate models for each property, which we refer to as STL-GNN. We train and evaluate each model 10 times with different seeds, and report average and standard deviation across these runs in all metrics.

3.3. Implementation & hyperparameters

The model is implemented using PyTorch and PyTorch Geometric (Fey and Lenssen, 2019) within our GMoLprop framework, which is available as open source. Hyperparameters are optimized with grid search, where the optimized parameters are the batch size ∈ {64, 128}, the activation function ∈ {LeakyReLU, SiLU} – the commonly used ReLU activation function is not considered, as it showed less suitable for thermodynamic prediction tasks in previous works (Rittig et al., 2023; Pavšek et al., 2025) –, the fingerprint dimension ∈ {64, 128} and for Clapeyron-GNN, the weighting factor of the Clapeyron loss ∈ {0.1, 0.5, 1}. This results in a batch size of 64, a fingerprint dimension of 64 and LeakyReLU as activation function for the STL-GNN, a batch size of 64, a fingerprint dimension of 64 and LeakyReLU as activation function for the MTL-GNN, and a batch size of 64, a fingerprint dimension of 128, LeakyReLU as activation function and a weighting factor of 0.1 for the Clapeyron-GNN. This value of the weighting factor yields the best performance as it does not negatively impact data approximation or lead to non-convergent training, while still achieving good approximation of the Clapeyron equation. Notably, the training performance is sensitive to the numeric value of the weighting factor, with larger values for λ resulting in non-convergent training. Including the weighting factor in the hyperparameter optimization is therefore essential. For the activation function, LeakyReLU results in better predictive performance than SiLU. While SiLU yields smooth functions with respect to temperature and lower Clapeyron errors, it significantly decreases predictive performance. Hence, we choose LeakyReLU as activation function.

4. Results & discussion

We first compare the Clapeyron-GNN to the other models, and then show exemplary predictions for individual molecules.

4.1. Model comparison

The overall prediction accuracy for all four properties across the three different models, STL-GNN, MTL-GNN and Clapeyron-GNN over 10 runs is shown in Table 1.

Comparing the performance of the MTL-GNN to the STL-GNNs, a clear increase in prediction performance can be observed for the vapor molar volume and the enthalpy of vaporization, for which data is rather scarce, see Section 3.1. For the liquid molar volume and the vapor pressure with relatively high data availability, the performance is comparable across the MTL-GNN and the STL-GNN. The relative improvement in the prediction performance of vapor molar volume and enthalpy of vaporization by the introduction of multi-task learning is

Table 1
Performance metrics, root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2) evaluated on the logarithmic scale on the test set for the three different models and for all four properties, and Clapeyron error (see Eq. (2)) evaluated on the test set. Large font values are averages and small fonts standard deviations.

Property	STL-GNN			MTL-GNN			Clapeyron-GNN		
	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
p^{sat}(T) (#78,840)	**0.27** ± 0.019	**0.14** ± 0.0074	**0.97** ± 0.0040	**0.26** ± 0.013	**0.14** ± 0.0076	**0.97** ± 0.0025	**0.26** ± 0.019	**0.14** ± 0.0065	**0.97** ± 0.0040
V^V(T) (#2206)	0.31 ± 0.028	0.18 ± 0.022	0.84 ± 0.028	**0.17** ± 0.022	**0.13** ± 0.021	**0.95** ± 0.013	**0.18** ± 0.019	**0.14** ± 0.016	**0.95** ± 0.011
V^L(T) (#43,056)	**0.048** ± 0.0057	**0.020** ± 0.0012	**0.95** ± 0.013	**0.048** ± 0.0029	0.030 ± 0.0022	**0.95** ± 0.0064	**0.048** ± 0.0050	0.029 ± 0.0046	**0.95** ± 0.012
ΔH_V(T) (#1057)	0.15 ± 0.024	0.099 ± 0.017	0.71 ± 0.096	**0.11** ± 0.016	**0.083** ± 0.017	**0.85** ± 0.046	**0.10** ± 0.023	**0.075** ± 0.018	**0.85** ± 0.063
$\mathcal{L}_{\text{Clapeyron}}$		0.45 ± 0.12			0.14 ± 0.050			**0.0069** ± 0.0051	

on par, i.e. the RMSE reduces from 0.31 to 0.17 and from 0.15 to 0.11, respectively. As the amount of data on vapor molar volume is twice as large as on the enthalpy of vaporization, a notable difference in the performance improvement through introduction of multi-task learning might have been expected. However, as this is not observed, small variations in dataset size appear not to have a significant influence on the value of multi-task learning. Performance improvements become observable at dataset size differences across orders of magnitude in our case. Hence in our case, multi-task learning introduces a clear advantage for properties where data are scarce, while for properties where data are abundant in the temperature domain, the prediction performance is not altered significantly. This aligns with the findings by bin Javaid et al. (2025), who find that MTL improves prediction accuracy only for tasks that have strong relationships among each other and limited data availability. Standley et al. (2020) report that in data scarce scenarios the change in predictive performance of MTL compared to STL depends largely on the combination of tasks that are trained jointly, significantly improving the performance for some and even worsening it for others, which can depend on the relationships between tasks. As in the present work, all four properties are strongly related — the existence of the Clapeyron equation serves as proof for that — our results align with literature reports with respect to the effect of MTL on predictive performance.

The prediction performance of the Clapeyron-GNN is on par with the high prediction performance of the MTL-GNN for all four prediction targets. However, in the Clapeyron error (see Table 1) a significant difference can be observed between the Clapeyron-GNN and the MTL-GNN. The Clapeyron error is with 0.007 compared to 0.138, respectively, two orders of magnitude smaller when using the thermodynamics-informed approach. Thus, for the single-species VLE prediction, Clapeyron regularization can significantly enhance the approximation of fundamental thermodynamics relations, without negatively impacting the performance in data approximation. The reason why it does not improve the prediction performance lies in the fact that the Clapeyron regularization provides additional information on the consistency of individual molecules. However, it does not give additional information on new molecules. As we test on unseen molecules, the prediction performance of the Clapeyron-GNN remains on the same level but rather shows improved consistency of predictions. Notably, also the introduction of MTL alone enhances adherence to the Clapeyron equation; for the STL-GNN, the Clapeyron error is significantly higher than for the MTL-GNN, with 0.45 compared to 0.14, respectively (see Table 1). So interestingly, the introduction of multi-task learning also improves the approximation of the Clapeyron equation, albeit significantly less than the introduction of Clapeyron regularization.

Overall, the Clapeyron-GNN achieves high prediction accuracy on the same level as MTL, while significantly better approximating fundamental thermodynamic relations. We thus further investigate the prediction capabilities of the Clapeyron-GNN for the best models and individual molecules in the following.

We show the parity plots of the best model out of the 10 different seeds, for all four properties for the MTL-GNN and the Clapeyron-GNN, see Fig. 3. We do not show the STL-GNN due to low accuracy. Overall, the best model of the Clapeyron-GNN slightly outperforms that of the MTL-GNN in two of the four properties (V^V(T), ΔH_V(T)). In all four properties, parallel lines to the diagonal are visible both for the predictions by the Clapeyron-GNN and by the MTL-GNN. This indicates that for individual molecules, the predictions follow the trend in the temperature dependency but contain an off-set. The groups of chemical compounds that contribute most strongly to this behavior are amines, alcohols and esters. For the vapor molar volume in the MTL-GNN case (see Fig. 3(c)), the clearly noticeable line above the diagonal between a real value of −4 and −2.5, which is almost perfectly parallel between a real value of −3.2 and −2.5, corresponds to trifluoromethane. Notably, other flourine compounds do not exhibit above average off-set in the predictions. As there is no clear trend visible in the compounds that yield off-sets in the predictions, a possible explanation is that the measurements themselves contain an off-set, although for trifluoromethane the data come from four different measurement series that are consistent among each other.

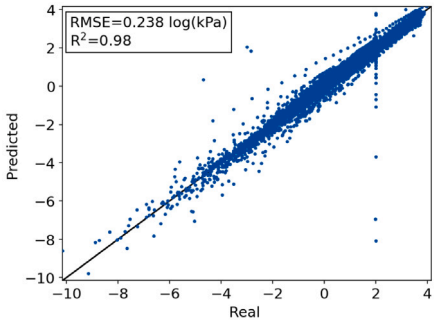
For the vapor molar volume and for the enthalpy of vaporization, see Fig. 3(c)–(d) &(g)–(h), the predictions of the Clapeyron-GNN align more closely aligned with the diagonal than for the MTL-GNN. However, both models overestimate the vapor molar volume in low molar volume regions (numeric values under −3.5). For most molecules, this region is close to the critical point, where predictions are inherently more difficult.

Overall — also for the best models of the MTL-GNN and the Clapeyron-GNN — the predictions are on a high level and capture the temperature dependency of the properties, with minor systematic deviations in both models on the verge of the respective data domains.

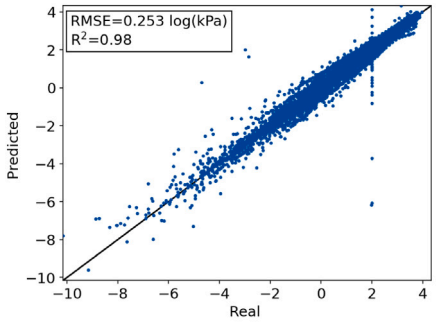
4.2. Predictions for individual molecules

In Fig. 4, we show the p(V)-plot and ΔH_v(T)-plot for predictions of the MTL-GNN (in blue) the Clapeyron-GNN (in green), and the Clapeyron-GNN with SiLU as activation function (in orange) for four molecules from the test set, namely 1-Bromobutane, 2-Bromopropane, 3-Methylthiophene and Piperidine. The four representative molecules are randomly selected from the subset of molecules in the test set that contain more than one datapoint for the enthalpy, to allow for comparison with experimental data.

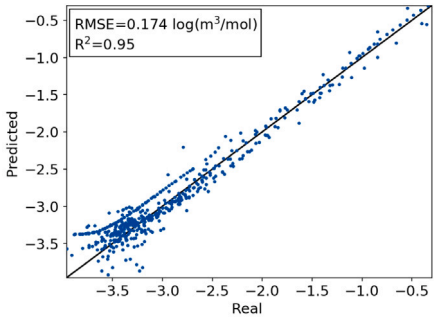
In the p(V)-plots, good alignment of both the Clapeyron-GNN and MTL-GNN predictions with the experimental data can be observed across orders of magnitude in the pressure and the molar volumes. In the lower pressure region, the vapor molar volume predictions deviate significantly between the MTL-GNN and the Clapeyron-GNN, especially for 1-Bromobutane 4(e). As in this area, there is no experimental data, it is not clear which prediction describes the true behavior more closely. For 1-Bromobutane and 2-Bromopropane, both



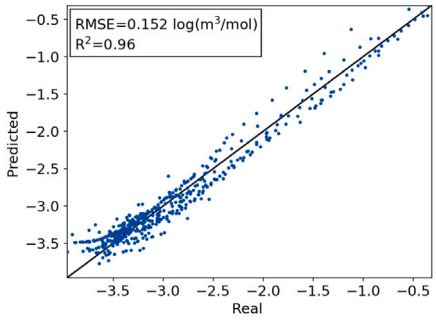
(a) Vapor pressure MTL-GNN



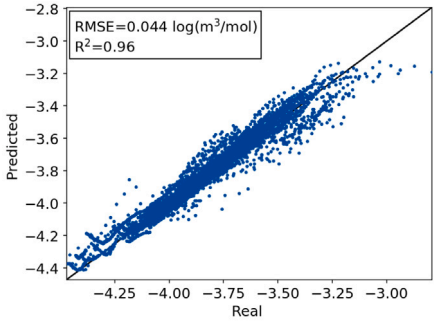
(b) Vapor pressure Clapeyron-GNN



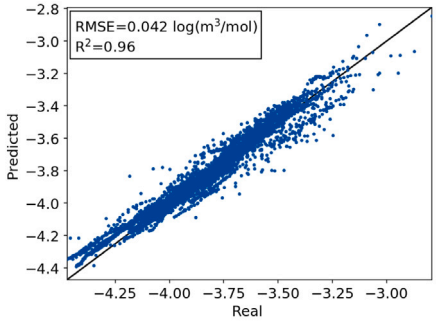
(c) Vapor molar volume MTL-GNN



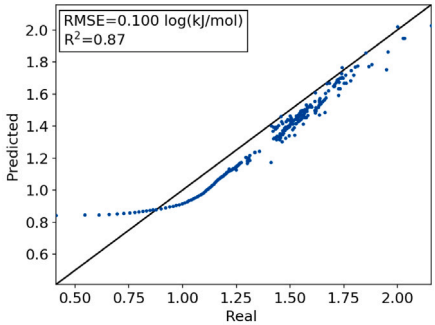
(d) Vapor molar volume Clapeyron-GNN



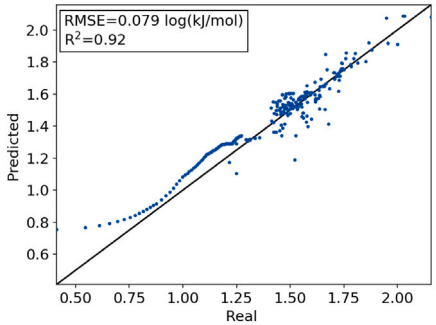
(e) Liquid molar volume MTL-GNN



(f) Liquid molar volume Clapeyron-GNN



(g) Enthalpy of vaporization MTL-GNN

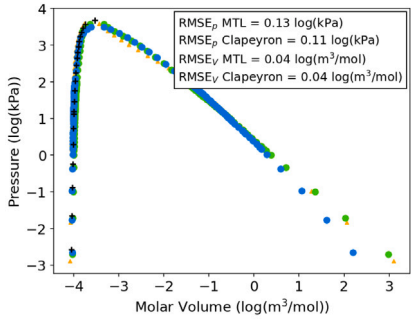


(h) Enthalpy of vaporization Clapeyron-GNN

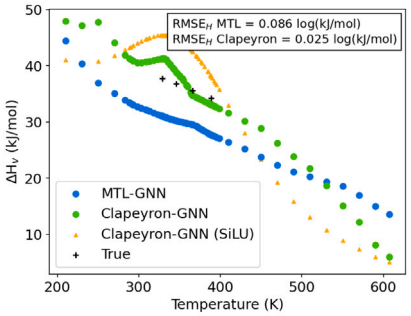
Fig. 3. Parity plots of test set MTL-GNN and Clapeyron-GNN.

the Clapeyron-GNN and the MTL-GNN approximate the critical point well. For 3-Methylthiophene both models underestimate the numeric values around the critical point, with the MTL-GNN underestimating more significantly, while for Piperidine the MTL-GNN overestimates the numeric values. Accurate predictions close to the critical point are inherently difficult, underlining the quality of the Clapeyron-GNN and MTL-GNN predictions, particularly as the models have not been trained on data from these specific molecules.

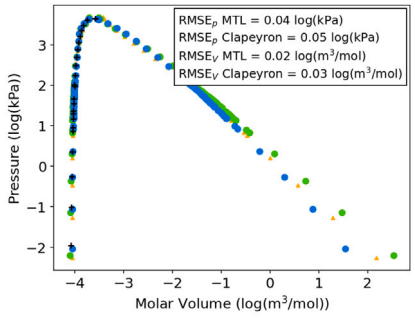
For the enthalpy, a significant difference in the prediction accuracy is observable. The Clapeyron-GNN approximates the data significantly better but exhibits a corner, while the predictions by the MTL-GNN have a constant off-set to the experimental data. This off-set is also



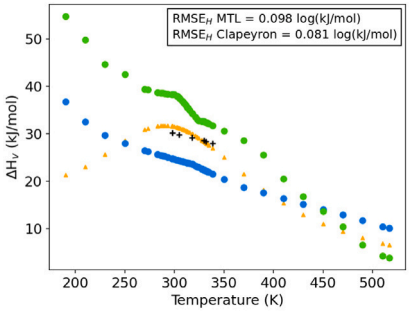
(a) 3-Methylthiophene



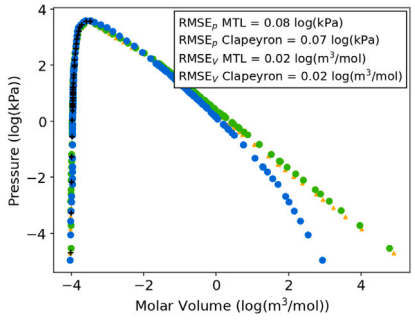
(b) 3-Methylthiophene



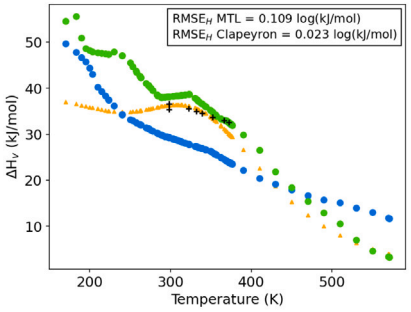
(c) 2-Bromopropane



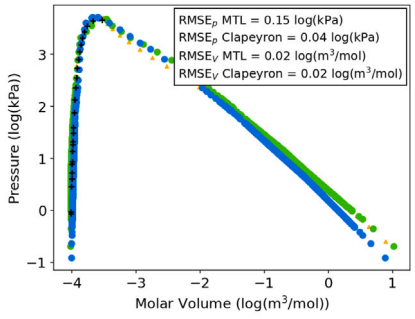
(d) 2-Bromopropane



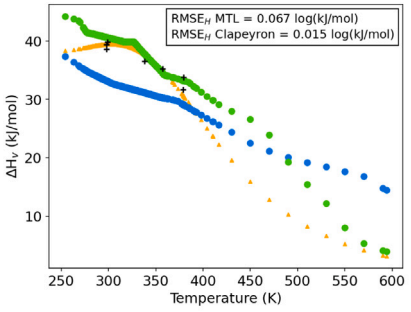
(e) 1-Bromobutane



(f) 1-Bromobutane



(g) Piperidine



(h) Piperidine

Fig. 4. p(V)-plots and ΔH_v(T)-plots for four exemplary molecules of the test set: experimental data in black crosses, multi-task learning in blue dots, Clapeyron-informed learning with LeakyReLU in green dots, and Clapeyron-informed learning with SiLU in orange triangles.

visible in the parity plot for enthalpy of vaporization of the MTL-GNN, where most prediction lie below the diagonal (see Section 4.1).

Apart from the approximation of the data, there is also a large difference in the trend of the MTL-GNN and the Clapeyron-GNN at higher temperatures. While the MTL-GNN predicts a linear trend with constant gradient with respect to temperature which seems close to the gradient in the experimental data, the Clapeyron-GNN predicts a steeper curve at higher temperatures, predicting significantly lower enthalpies at high temperatures than the MTL-GNN. These predictions are more consistent as the enthalpy of vaporization approaches zero, as the temperature approaches the critical temperature. Given the fact, that there is little data on the enthalpy across the temperature domain

for most molecules, learning this trend from data alone is challenging. This is where the strength of the Clapeyron-regularization can be observed, as it improves the consistency of predictions particularly in these areas close to the critical point where data are scarce, and the purely data-driven MTL-GNN fails to make consistent predictions.

However, flaws can also be observed in the Clapeyron-GNN predictions. At medium to lower temperatures corner points are visible in the Clapeyron-GNN predictions. This behavior is non-physical as the enthalpy of vaporization decreases in a smooth curve from its maximum at the triple point until it approaches zero at the critical point. Notably, when employing SiLU as activation function, the Clapeyron-GNN yields smooth output functions, but has a local maximum in the enthalpy for all four example molecules. While there are corners in the prediction of the enthalpy when using LeakyReLU as activation function for the Clapeyron-GNN, it does follow the overall trend of the data better, even beyond the temperature range in the training data. Hence, LeakyReLU is the better choice for balancing accuracy and consistency (see Section 3.3). As data are scarce for the enthalpy, it is likely that the loss signal during training is dominated by the Clapeyron regularization, which might introduce corner points when there is inconsistencies in the experimental data between the four properties. This underlines that thermodynamics-informed models only promote but do not guarantee consistency and may be susceptible to noise in the training data, particularly in data-scarce settings. If data were sufficiently dense across the temperature domain for all four properties, the minimization of the data approximation error would make such non-physical predictions unlikely. However, especially when data are scarce, non-physical predictions are possible.

Nonetheless, Clapeyron-GNN shows highly promising performance, enabling good approximation of Clapeyron-consistent VLE calculation for single-species systems, for which experimental data is lacking for conventional VLE calculations, e.g. via Peng Robinson (Robinson et al., 1985).

5. Conclusion

We incorporate the Clapeyron equation into the training of GNNs for predicting single-species vapor–liquid equilibria from molecular graph structure and temperature in a multi-task learning setting. We find Clapeyron regularization of model training to greatly increase the approximation of the exact Clapeyron equation while maintaining the same level of prediction accuracy. Additionally, multi-task learning of vapor molar volume, liquid molar volume, vapor pressure and enthalpy of vaporization significantly increases prediction accuracy compared to single-task learning. Clapeyron-GNN is therefore a highly promising architecture for single-species VLE predictions for molecular and process design applications. Beyond the specific application, multi-task learning in combination with physics-enhanced approaches is a promising approach for thermodynamic prediction tasks in general, where there are multiple correlated properties of interest.

Future work should investigate the possibility of using the Clapeyron equation as a hard constraint in the model architecture, following our recently introduced thermodynamics-consistent GNN approach (Rittig and Mitsos, 2024), i.e., embedding the Clapeyron equation directly into the output head of the model. This would be particularly interesting, as the present results, in particular for the enthalpy, show that employing physics-based regularization does not guarantee that all predictions are physically consistent in all scenarios. Notably, our first investigations of a thermodynamics-consistent GNN with embedded Clapeyron equation did result in lower prediction accuracy, requiring further tests to be performed in future work, targeting a wider range of properties and molecules. For extending data sets and model testing in practical scenarios, collaboration with industry thus remains critical.

CRediT authorship contribution statement

Jan Pavšek: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Alexander Mitsos:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Elvis J. Sim:** Writing – review & editing, Software, Methodology, Investigation. **Jan G. Rittig:** Writing – review & editing, Supervision, Software, Conceptualization.

Software availability

The code is available in our *GitLab* repository GMoLprop.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), Germany – 466417970 – within the Priority Programme “SPP 2331: Machine Learning in Chemical Engineering”.

Model training was performed with computing resources granted by RWTH Aachen University, Germany.

The authors thank René Görgen for his software engineering support.

Data availability

The data used for training of the presented models are confidential and can be accessed through the NIST Thermodata engine.

NIST Thermodata engine

References

Alam, M.Z., Alqallaf, A., McGill, C., 2026. Equinet: Modeling vapor–liquid equilibrium using simultaneous neural-network predictions for activity coefficient and vapor pressure.

bin Javaid, M., Gervens, T., Mitsos, A., Grohe, M., Rittig, J.G., 2025. Exploring data augmentation: Multi-task methods for molecular property prediction. Comput. Chem. Eng. 109253.

Fey, M., Lenssen, J.E., 2019. Fast graph representation learning with PyTorch Geometric. In: ICLR Workshop on Representation Learning on Graphs and Manifolds.

Heid, E., Greenman, K.P., Chung, Y., Li, S.-C., Graff, D.E., Vermeire, F.H., Wu, H., Green, W.H., McGill, C.J., 2023. Chemprop: a machine learning package for chemical property prediction. J. Chem. Inf. Model. 64 (1), 9–17.

Kochi, M.R., Rezaei, H., Khan, S.T., Mamillapalli, B.T., Ebrahimiazar, M., Ye, H., Moosavian, R., Zargartalebi, M., Sinton, D., Moosavi, S.M., 2025. Thermodynamics-informed machine learning for predicting temperature-dependent chemical properties.

Leenhouts, R., Jankelevitch, S., Raike, R., Müller, S., Vermeire, F., 2025. Thermodynamics-informed graph neural networks for phase transition enthalpies. Syst. Control. Trans. 1662–1669.

Liu, S., Qu, M., Zhang, Z., Cai, H., Tang, J., 2021. Multi-task learning with domain knowledge for molecular property prediction. In: NeurIPS 2021 AI for Science Workshop.

Medina, E.I.S., Sundmacher, K., 2026. Graph neural networks embedded into margules model for vapor–liquid equilibria prediction. Fluid Phase Equilib. 599, 114520.

National Institute of Standards and Technology (NIST), 2025. Thermodata engine, nist standard reference database 103b, version 10.4.5. critically evaluated thermodynamic and transport property data. URL: https://www.nist.gov/mml/acmd/trc/thermodata-engine/srd-nist-tde-103b. (Accessed 20 June 2025).

Park, J., Muthoka, R.M., Jang, S., Lee, Y., 2025. A multi-stage graph neural network–physics-informed neural network (gnn–pinn) framework for thermodynamic property prediction. Ind. Eng. Chem. Res. 64 (40), 19722–19734.

Pavšek, J., Mitsos, A., Dahmen, M., Tan, T.X., Rittig, J.G., 2025. Deepeosnet: Capturing the dependency on thermodynamic state in property prediction tasks. arXiv preprint arXiv:2509.17018.

Qin, S., Jiang, S., Li, J., Balaprakash, P., Van Lehn, R.C., Zavala, V.M., 2023. Capturing molecular interactions in graph neural networks: a case study in multi-component phase equilibrium. Digit. Discov. 2 (1), 138–151.

Raissi, M., Perdikaris, P., Karniadakis, G.E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J. Comput. Phys. 378, 686–707.

Rittig, J.G., Dahmen, M., Grohe, M., Schwaller, P., Mitsos, A., 2026. Molecular machine learning in chemical process design. Curr. Opin. Chem. Eng. 52, 101239.

Rittig, J.G., Felton, K.C., Lapkin, A.A., Mitsos, A., 2023. Gibbs–duhem-informed neural networks for binary activity coefficient prediction. Digit. Discov. 2 (6), 1752–1767.

Rittig, J.G., Mitsos, A., 2024. Thermodynamics-consistent graph neural networks. Chem. Sci. 15 (44), 18504–18512.

Robinson, D.B., Peng, D.-Y., Chung, S.Y., 1985. The development of the peng-robinson equation and its application to phase equilibrium in a system containing methanol. Fluid Phase Equilib. 24 (1–2), 25–41.

Rosenberger, D., Barros, K., Germann, T.C., Lubbers, N., 2022. Machine learning of consistent thermodynamic models using automatic differentiation. Phys. Rev. E 105 (4), 045301.

Schweidtmann, A.M., Rittig, J.G., Konig, A., Grohe, M., Mitsos, A., Dahmen, M., 2020. Graph neural networks for prediction of fuel ignition quality. Energy Fuels 34 (9), 11395–11407.

Sosnin, S., Vashurina, M., Withnall, M., Karpov, P., Fedorov, M., Tetko, I.V., 2019. A survey of multi-task learning methods in chemoinformatics. Mol. Inform. 38 (4), 1800108.

Specht, T., Nagda, M., Fellenz, S., Mandt, S., Hasse, H., Jirasek, F., 2024. Hanna: hard-constraint neural network for consistent activity coefficient prediction. Chem. Sci. 15 (47), 19777–19786.

Standley, T., Zamir, A., Chen, D., Guibas, L., Malik, J., Savarese, S., 2020. Which tasks should be learned together in multi-task learning? In: International Conference on Machine Learning. PMLR, pp. 9120–9132.

Wahyudi, A., Sueviriyapan, N., Rirksomboon, T., Suriyapraphadilok, U., et al., 2026. Deepthermomix: a local composition graph neural networks model for multicomponent activity coefficients.

Winter, B., Rehner, P., Esper, T., Schilling, J., Bardow, A., 2025. Understanding the language of molecules: predicting pure component parameters for the pc-saft equation of state from smiles. Digit. Discov. 4 (5), 1142–1157.

Wu, Z., Wang, H., He, C., Zhang, B., Xu, T., Chen, Q., 2023. The application of physics-informed machine learning in multiphysics modeling in chemical engineering. Ind. Eng. Chem. Res. 62 (44), 18178–18204.