Systems & Control Transactions

Research Article - Peer Reviewed Conference Proceeding
ESCAPE 35 - European Symposium on Computer Aided Process Engineering
Ghent, Belgium. 6-9 July 2025
Jan F.M. Van Impe, Grégoire Léonard, Satyajeet S. Bhonsale,
Monika E. Polańska, Filip Logist (Eds.)

PSE PRESS

Thermodynamics-informed Graph Neural Networks for Phase Transition Enthalpies

Roel Leenhouts[a], Sebastien Jankelevitch[a], Roel Raike[a], Simon Müller[b] and Florence Vermeire[a]*

[a] Department of Chemical Engineering, KU Leuven, Leuven, Belgium
[b] Institute of Thermal Separation Processes, Hamburg University of Technology, Hamburg, Germany
* Corresponding Author: florence.vermeire@kuleuven.be.

ABSTRACT

Phase transition enthalpies, such as those for fusion, vaporization, and sublimation, are vital for understanding thermodynamic properties and aiding early-stage process design. However, measuring these properties is often time-consuming and costly, leading to increased interest in computational methods for fast and accurate predictions. Graph neural networks (GNNs), known for their ability to learn complex molecular representations, have emerged as state-of-the-art tools for predicting various thermophysical properties. Despite their success, GNNs do not inherently obey thermodynamic laws. In this study, we present a multitask GNN designed to predict vaporization, fusion, and sublimation enthalpies of organic compounds. We modified the loss function of the GNN, accounting for the thermodynamic cycle of the three phase transition enthalpies. To train the model, we digitized the extensive Chickos and Acree compendium, which encompasses 32,023 experimental measurements. Two approaches were explored: soft constraints, which guide the model toward thermodynamic consistency, and hard constraints, which enforce fully consistent predictions. The GNN achieved root mean squared errors (RMSEs) of 19.9 kJ/mol for sublimation, 11.0 kJ/mol for fusion, and 16.5 kJ/mol for vaporization enthalpies on the test set. Soft constraints were found to provide a good balance between accuracy and thermodynamic consistency, whereas hard constraints prioritized fidelity at the expense of predictive performance. When compared to the conventional Joback group contribution method the GNN demonstrated an improved accuracy and applicability range. This work underscores the potential of thermodynamics-informed GNNs for predicting thermodynamic properties accurately while maintaining consistency, paving the way for more reliable and efficient computational approaches.

Keywords: Graph neural networks, Property prediction, Physics informed, Phase transition enthalpies

INTRODUCTION

Phase transition enthalpies, such as sublimation, melting, and vaporization enthalpies, have numerous applications for designing process unit operations. For example, they are a crucial part of the energy balances of separation units such as distillation towers or crystallization units. Additionally, phase transition enthalpies can be applied within thermodynamic cycles to predict different thermochemical properties[1],[2]. Measuring these properties is time intensive and costly, therefore it is important to develop predictive methods for fast screening of potential materials.

Different approaches have been used to predict these properties. First principal methods are accurate, in particular for vaporization enthalpies[3], but they are computationally expensive and fall short in crystal structure prediction. On the other hand, empirical methods, such as quantitative structure property relationships (QSPR) [4] and the linear solvation energy relationships (LSER) [5], are commonly used. They can be accurate but have a limited applicability range. Especially for novel components, these empirical methods often lack accuracy. Recent improvements in data availability and computational resources drive the use of machine learning to develop robust and fast models to predict phase transition enthalpies. Graph neural networks have become the state-of-the-art method for molecular property prediction because of their flexibility in molecule representation and their ability to learn deep representations [6]. They have consistently surpassed fixed molecular representation methods [7] and are therefore utilized for analyzing

the properties explored in this study.

One limitation of machine learning algorithms is that they do not inherently imply thermodynamic consistency. Recent studies have investigated the inclusion of these thermodynamic laws in the loss function of deep learning property prediction methods [8],[9] to improve the confidence in black-box models. For phase transition enthalpies, we consider thermodynamic consistency in our models by implying the following:

$$\Delta H_{sub,T} = \Delta H_{fus,T} + \Delta H_{vap,T} \quad (1)$$

Two approaches are reported for the inclusion of this law in the loss function. In the first approach, a regularization term is added to the loss function as a soft constraint, where inconsistent predictions are still allowed but penalized. In the second approach, the loss function is reformulated to only allow consistent predictions or also called hard constraint.

To train our models, we digitized the compendium of Chickos and Acree[10],[11],[12], consisting of 32,023 phase transition enthalpy values measured over 135 years. The performance of the machine learning methods is compared to the Joback group contribution method as implemented in the thermo python package[4]. Additionally, the performance is evaluated by comparing single-task and multi-task models, as well as soft and hard constraint loss functions.

METHODS

Phase Transition Enthalpy Database

Database digitization

The data collected from the works of Chickos and Acree's enthalpy compendium[10],[11],[12] is detailed here. In the compendium for each data point the following metrics were reported: the molecular formula, the CAS-number, the measurement temperature, the enthalpy value, the measurement technique, and the original source. From the CAS-numbers, the InChI and SMILES identifiers were retrieved from the PubChem database[13] and Cactus service[14]. Molecules that include metal atoms were excluded, so the dataset is limited to the following atoms list: C, H, B, Br, Cl, F, I, N, O, P, S, and Si.

Additionally, the enthalpies were converted to the standard state (298 K, 1 atm) using the approach proposed by Chickos et al.[15]. This correction was required, since Equation 1 is only applicable when the three phase enthalpies are at the same temperature. For this correction, heat capacity values were required for the solid, liquid and gas phase. The heat capacities of the solid and liquid phase were computed by the group contribution method from Chickos et al.[15], and for the gas phase, temperature dependent heat capacities were calculated using the RMG software[16]. A list of the heat capacities is provided as part of our Research Data Repository.

Database characteristics

Analyzing the dataset, Figures 1 and 2 show the distributions of the enthalpy values and molecular weights respectively for the three phase transition enthalpies. The reported fusion enthalpy values were smaller in magnitude and distribution width, which is expected as, compared to the vaporization and sublimation enthalpy, they do not depend on the (higher) enthalpy difference with the gas phase.

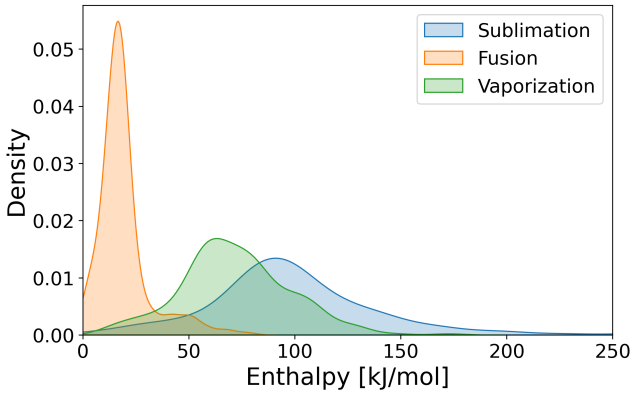

Figure 1. Value distribution of the three phase transition enthalpies at 298 K.

The molecular weight distributions for all three enthalpies are showing a widespread, which is critical for training a widely applicable machine learning model. The fusion enthalpy distribution is slightly more shifted to the right, which is inherent for these properties considering organic compounds. For many organic molecules with higher molecular weights, it is not possible to measure vaporization or sublimation enthalpies, since the decomposition temperature will be lower than the boiling point.

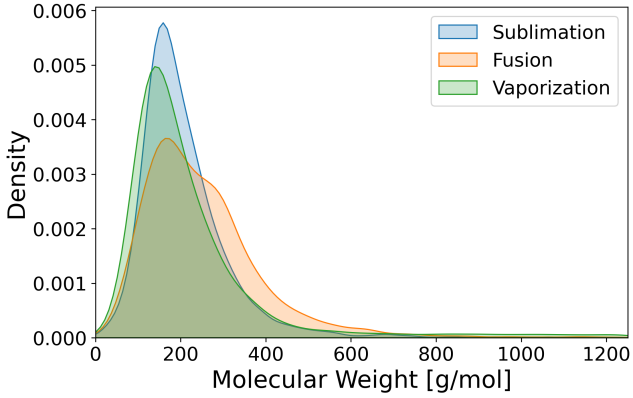

Figure 2. Molecular weight distribution of the three phase transition enthalpies.

For data points where all three enthalpies were available, the sum of the fusion and vaporization

enthalpies was compared to the sublimation enthalpy. In Figure 3, a parity plot for this thermodynamic consistency check of the database is shown. It was found that, although the properties are highly correlated, large deviations from the thermodynamic cycle were apparent. It was observed that the experimental sublimation enthalpy is almost consitently higher as compared to the one calculated from the vaporization and fusion enthalpy, with some deviations higher than 50 kJ/mol.

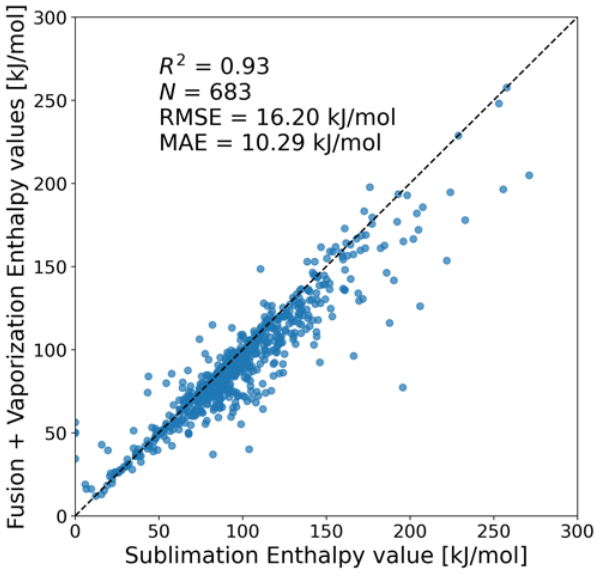

Figure 3. Thermodynamics consistency parity plot for the overlapping values in the phase transition enthalpies at 298 K of the database.

Crystal structures were not reported in the compendium, and different polymorphs could partially account for the observed deviations. Additionally, temperature corrections introduce uncertainty; however, they have been shown to produce acceptable errors and are essential for assessing thermodynamic consistency [11]. The temperature correction for the enthalpy of vaporization introduces an error of 0.016 kJ/(mol K) [11], leading to an uncertainty of 1.6 kJ/mol for a molecule deviating by 100 K from the reference temperature of 298 K. For enthalpies of sublimation and vaporization, an uncertainty of one-third of the total temperature correction can be assumed [11]. As an example, for carbazole, this results in an enthalpy of sublimation of 102.9 ± 1.2 kJ/mol and an enthalpy of fusion of 15.9 ± 3.5 kJ/mol.

The average estimated uncertainties for the temperature correction over the dataset were 0.57, 2.38, and 1.32 kJ/mol for sublimation, fusion, and vaporization enthalpies, respectively. Since sublimation and vaporization enthalpies are commonly reported in the literature at values corrected to 298 K, these uncertainties were averaged only for values that were not reported at 298 K.

Figure 4 shows the distribution of standard deviations for duplicated data points in the dataset before the temperature correction. Duplicates were identified for 1,312 sublimation, 1,143 fusion, and 2,663 vaporization compounds. These standard deviations provide an estimate of experimental uncertainty, specifically aleatoric uncertainty. The average standard deviations are 5.9, 3.3, and 5.4 kJ/mol for sublimation, fusion, and vaporization enthalpies, respectively. This represents the performance limit for a model trained on this data. Notably, the uncertainties introduced by temperature corrections are relatively small compared to experimental uncertainty.

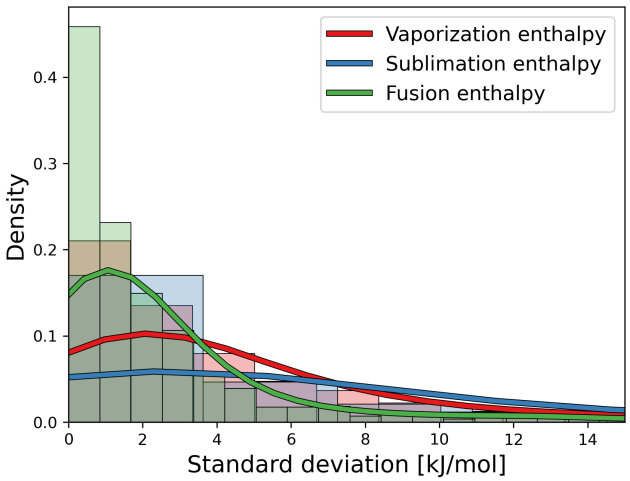

Figure 4. A distribution of the standard deviations computed on the duplicated data points in the dataset before the data was adjusted to 298 K.

Model architecture and training

The graph neural networks (GNNs) were trained using the SolProp repository [2], which focuses on solvation properties and implements directed message passing neural networks (D-MPNN) based on the Chemprop framework [6]. For a detailed explanation, readers are referred to dedicated studies [6],[17]. Figure 5 presents an example model architecture. The model takes a molecular identifier, such as SMILES or InChI, as input and converts it into a featurized molecular graph. This graph representation is then processed by the GNN, which predicts the enthalpy values for phase transitions.

Four different model architectures were trained: single-task, multitask, soft-constraint, and hard-constraint models. The single-task models were trained by initializing three separate neural networks, each focusing on one of the three enthalpy datasets (fusion, vaporization, and sublimation). In contrast, the multitask models used a single GNN with a multi-head output for the multi-layer perceptron (MLP), meaning that a single molecular graph was provided as input, and the network simultaneously predicted all three enthalpies.

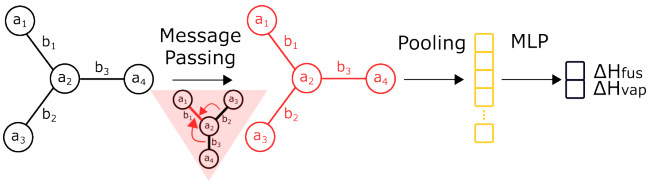

Figure 5. Hard-constraint thermodynamics-informed architecture.

The soft-constraint models used a multi-head output for the MLP, but the loss function was modified to include a penalty term that encouraged thermodynamic consistency. This penalty term was scaled by a weight parameter, λ, to balance accuracy in minimizing mean squared error (MSE) while ensuring the enthalpies remained consistent with thermodynamic relationships. After evaluating three values (0.1, 0.01, and 0.001), a λ of 0.01 was found to significantly influence the penalty term without negatively affecting model accuracy. The loss function for soft-constraint models was defined as:

$$\text{Loss} = \sum_{i=1}^{3}\left(y_{i,\text{exp}} - \hat{y}_{i,\text{pred}}\right)^2 + \lambda\left(\Delta\hat{H}_{\text{sub,pred}} - \left(\Delta\hat{H}_{\text{fus,pred}} + \Delta\hat{H}_{\text{vap}}\right)\right)^2 \quad (2)$$

For the hard-constraint models, thermodynamic consistency was strictly enforced within the loss function. Instead of predicting all three enthalpies separately, a single neural network was trained with two outputs: one for fusion enthalpy and one for vaporization enthalpy. The sublimation enthalpy was not directly predicted but was instead computed within the loss function as the sum of the fusion and vaporization enthalpies. The loss function ensured consistency by penalizing deviations from this relationship:

$$\text{Loss} = \left(y_{\text{fus,exp}} - \hat{y}_{\text{fus,pred}}\right)^2 + \left(y_{\text{vap,exp}} - \hat{y}_{\text{vap,pred}}\right)^2 + \left(y_{\text{sub,exp}} - \hat{y}_{\text{fus,pred}} - \hat{y}_{\text{vap,pred}}\right)^2 \quad (3)$$

The dataset was randomly split into a training and validation set (90%) and a test set (10%). Within the training set, a ten-fold cross-validation ensemble was trained to improve model robustness. The standard deviations reported in the results stem from variations between the predictions of the ensemble models. The featurization and hyperparameters remained fixed across all models and were based on prior research in thermodynamic property prediction [17].

Joback method implementation

The Joback method[4] is a group contribution method that was developed for the prediction of a range of properties among others the vaporization and fusion enthalpy. Here, the implementation from the thermo v0.4.0 python package was used[18]. Predictions were made on the same test set as for the graph neural network. However, the fragmentation algorithm was not able to identify all atoms in the molecules. For the molecules where this was not possible, the predictions were removed from the results, accounting for 16.7% of the compounds in the test set.

Additionally, the Joback method predicts the phase transition enthalpies at the transition temperature, contrary to our method that predicts at the enthalpies at 298 K. Therefore, the fusion and vaporization enthalpies determined with the Joback method, were converted in the same way as the experimental dataset using the approach proposed by Chickos et al.[15]. Since the Joback method was not parametrized for the sublimation enthalpy, this was calculated by the sum of the converted vaporization and fusion enthalpies.

RESULTS AND DISCUSSION

Model performance

The model performance is compared for the four different model architectures in Table 1-3 for sublimation, fusion, and vaporization enthalpies respectively. In general, the fusion and vaporization enthalpy were predicted more accurately as compared to the sublimation enthalpy. The best models report an MAE of 13 kJ/mol, 6 kJ/mol, and 8 kJ/mol for the sublimation, fusion, and vaporization enthalpies in the test set respectively.

A similar performance was obtained between the single- and multitask models. A slightly higher RMSE and MAE can be observed for the predictions of the single task models, in particular for sublimation enthalpy. However, these differences are well within the standard deviations between the model ensembles (σ in Table 1-3) which is a measure for the variance contribution to the epistemic uncertainty.

Applying constraints to the multitask model did not lead to a significant improvement in accuracy. Introducing the soft constraint resulted in RMSE and MAE values for all three phase transition enthalpies that remained similar to those of the multitask model. However, when the hard constraint was applied, prediction accuracy declines. Specifically, the RMSE and MAE for the sublimation enthalpy—directly influenced by the hard constraint—increase by more than 10 kJ/mol. This suggests that enforcing thermodynamic consistency through a hard constraint reduces the accuracy of phase transition enthalpy predictions. Moreover, further research is needed to explore the most effective methods for incorporating constraints into neural network training.

Due to the correlation between the three phase transition enthalpies, one would expect a multitask modeling architecture to offer more benefits compared to a single-task model. Additionally, incorporating thermodynamic constraints into the loss function should ideally lead to more accurate predictions. However, as previously discussed, no significant improvements in model accuracy were observed. When comparing model

performance to the average standard deviation from the repeated experiments, the MAE for the fusion and sublimation enthalpies was roughly twice the standard deviation found in the Database Characteristics section. In contrast, the MAE for the vaporization enthalpy was only about 2 kJ/mol from the standard deviation in the dataset, suggesting that the model is close to the aleatoric limit.

A potential explanation for the lower performance of the sublimation and fusion models could be the absence of crystal structure information. The repeated data points in the compendium often came from the same sources or authors, meaning the compounds—and likely their crystal structures—were quite similar. This may have resulted in a low estimate of aleatoric uncertainty. However, further investigation is needed to reach a definitive conclusion.

It is also worth noting that the MAE for the thermodynamic consistency of the experimental data (Figure 3) is 10.3 kJ/mol. While this is not a direct measure of experimental uncertainty, it could offer an indication of its magnitude. Importantly, the MAE for the phase transition enthalpy predictions is lower than 10.3 kJ/mol, except for the sublimation enthalpy.

Lin et al. achieved a RMSE of 4.8 kJ/mol for vaporization enthalpies using first-principles calculations [19], which is notably lower than the standard deviation observed in our data. While our model demonstrates lower accuracy, direct comparison is challenging due to differences in methodologies and test sets. For fusion enthalpies, Chickos et al. reported a prediction error of 6.4 kJ/mol [20], closely matching our model's accuracy. Similarly, Liu et al. used first-principles calculations to estimate sublimation enthalpies, obtaining a mean absolute error (MAE) of 13.5 kJ/mol compared to experimental values [21], which aligns with our model's performance.

Table 1: An overview of the model architectures performance for the sublimation enthalpy on the test set.

Model	RMSE (kJ/mol)	MAE (kJ/mol)	σ (kJ/mol)
Single task	20.2	13.9	5.1
Multitask	19.9	13.1	3.7
Soft constraint	19.9	13.1	1.4
Hard constraint	30.1	24.1	2.8

Table 2: An overview of the model architectures performance for the fusion enthalpy on the test set.

Model	RMSE (kJ/mol)	MAE (kJ/mol)	σ (kJ/mol)
Single task	11.0	6.3	2.1
Multitask	11.0	6.1	1.5
Soft constraint	11.0	6.3	3.5
Hard constraint	11.6	6.9	1.4

Table 3: An overview of the model architectures performance for the vaporization enthalpy on the test set.

Model	RMSE (kJ/mol)	MAE (kJ/mol)	σ (kJ/mol)
Single task	16.2	7.3	2.8
Multitask	16.2	7.5	3.1
Soft constraint	16.5	7.5	2.9
Hard constraint	17.5	8.9	2.0

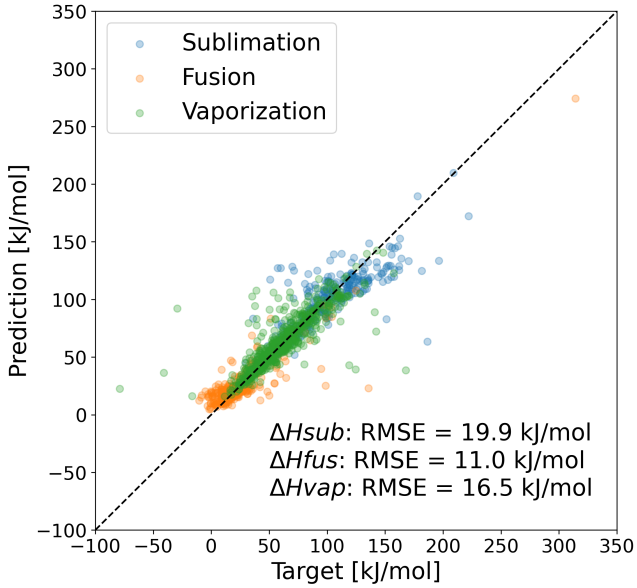

Figure 6. Parity plot for the enthalpy predictions made by the SolProp soft-constraint model.

The model predictions for the three phase transition enthalpies are visualized in Figure 6 for the model that includes the soft constraint in the loss function. The majority of the test data is predicted around the parity line, while large deviations are observed for some specific datapoints. The highest deviations for vaporization enthalpy are observed for negative experimental vaporization enthalpies. There were a small group of high molecular weight hydrocarbons for which the temperature correction resulted in a negative value because of the high measurement temperature compared to the reference temperature. Due to the large temperature range the temperature correction is expected to be large too. The low amount of available data for the heavy hydrocarbons and the high uncertainty on the correct can explain the lower performance of the models on this data. We did not find a correlation between the highest errors and the experimental data that showed the highest deviations from thermodynamic consistency in Figure 3, however only 5% of the test set contained experimental measurements for all three phase transition enthalpies. In addition, there was no strong correlation between the epistemic model uncertainties from the model ensemble and the

prediction error.

Thermodynamic consistency

On top of the model performance, the thermodynamic consistency of the predictions was evaluated. In Figure 7, the predicted sublimation enthalpy is compared to the sublimation enthalpy calculated from the fusion and vaporization enthalpy through Equation 1. The thermodynamic consistency is visualized for the multitask model as well as the model with the soft constraint. The thermodynamic consistency for the model with the hard constraint is omitted from Figure 7 since, as expected from the constraint, the calculated sublimation enthalpy is exactly equal to the predicted sublimation enthalpy.

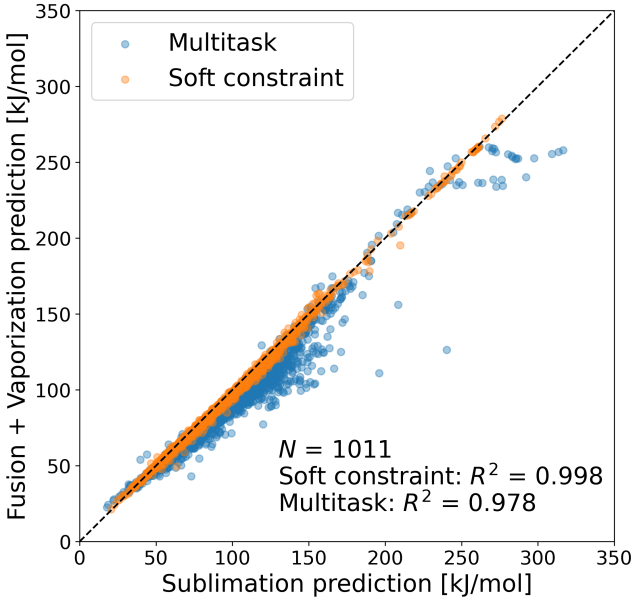

Figure 7. Parity plot for the thermodynamic consistency of the multitask and soft constraint model.

The multitask modeling approach already improved the thermodynamic consistency as compared to the experimental measurements with overlapping phase transition enthalpies (**Figure 3**). Yet, the calculated sublimation enthalpy was consistently lower compared to the sublimation enthalpy predicted by the multitask model. The model indeed learned the consistent underestimations that are present in the experimental database. Applying the soft constraint corrects for these discrepancies and a good agreement between calculated and predicted sublimation enthalpy was obtained (R^2=0.998). Interestingly, the performance of the model is hardly affected by applying this constraint while the predictions are more consistent. The soft constraint therefore provides a good balance between model performance and consistent predictions. Note that increasing the value of λ in the loss function would result in a higher thermodynamic consistency but at the cost of a lower accuracy of the model.

Comparison to conventional methods

At last, the performance of our models was compared to the Joback method, which is the conventional way of determining phase transition enthalpies. In Figure 8, a parity plot is given comparing the experimental phase transition enthalpies with the Joback calculations as implemented in the thermo python library. Note that the sublimation enthalpy was determined through the thermodynamic cycle from Equation 1 as it is not directly available from the Joback method.

For fusion and vaporization enthalpies respectively, the RMSEs is the Joback method are 21.8 kJ/mol and 19.4 kJ/mol as compared to 11.0 kJ/mol and 16.5 kJ/mol from our soft constraint model. As such, it is demonstrated that advancements in data availability and computational algorithms can provide a significant improvement in property prediction as compared to conventional methods.

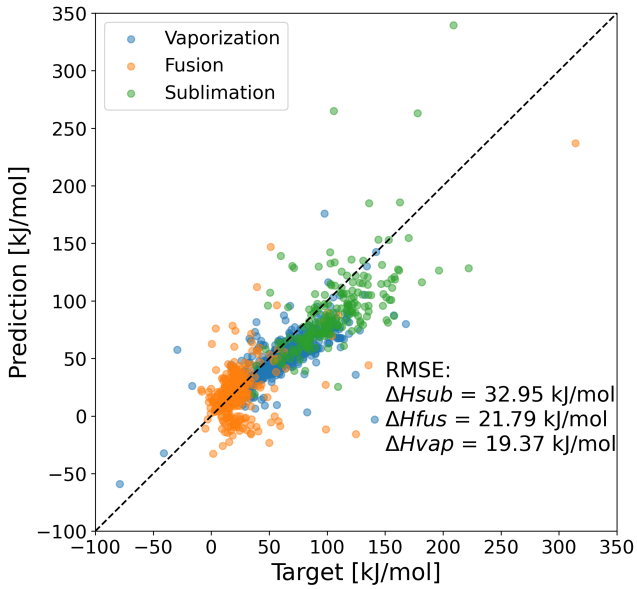

Figure 8. Parity plot for the enthalpy predictions made by the Joback method adjusted to standard state[4],[18].

CONCLUSIONS

In this work, we digitized the compendium of Chickos and Acree for the phase transition enthalpies of sublimation, fusion, and vaporization. The availability of such large and diverse database stimulates the development of deep learning algorithms for the training and prediction of these enthalpies.

Graph neural networks are used to predict these phase transition enthalpies while accounting for thermodynamics. First, no significant improvement in accuracy was found whilst comparing single- and multitask modeling architectures. Additionally, adding a soft-constraint thermodynamics loss function proved to be the best

balance between accuracy and thermodynamic consistency of the predictions. The hard-constraint loss function enforced strict consistency, however at the cost of the accuracy of the model specifically for sublimation enthalpies. A comparison to the Joback method showed an improvement in accuracy and applicability range of the model.

Our graph neural network achieved state-of-the-art performance on sublimation, fusion, and vaporization enthalpies, while adding thermodynamic consistency. This method has the potential to be applied to the screening materials in the early design phase of chemical processes for all three investigated phase transition enthalpies.

DIGITAL SUPPLEMENTARY MATERIAL

The code repository of the work can be found at the KU Leuven GitLab:

https://gitlab.kuleuven.be/creas/vermeire-group/solprop/-/tree/phase_transition_enthalpies,

and the digitized dataset and static code at KU Leuven Research Data Repository (RDR):

https://doi.org/10.48804/CBHEAB.

ACKNOWLEDGEMENTS

The authors want to thank W. Acree and J. Chickos for sharing a pdf of their compendium. We acknowledge the Fonds Wetenschappelijk Onderzoek (FWO) for funding (G021924N).

REFERENCES

1. Zhu, X. , Polyakov, V. R. , Bajjuri, K. , Hu, H. , Maderna, A. , Tovee, C. A. , and Ward, S. C. Building Machine Learning Small Molecule Melting Points and Solubility Models Using CCDC Melting Points Dataset. Journal of Chemical Information and Modeling 2023, 63, 2948–2959.
2. Vermeire, F. H. , Chung, Y. , and Green, W. H. Predicting Solubility Limits of Organic Solutes for a Wide Range of Solvents and Temperatures. Journal of the American Chemical Society 2022, 144, 10785–10797.
3. Lin, S. T. , Chang, J. , Wang, S. , Goddard, W. A. , and Sandler, S. I. Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model. Journal of Physical Chemistry A 2004, 108, 7429–7439.
4. [Joback, K. G. A unified approach to physical property estimation using multivariate statistical techniques. 1984
5. Abraham, M. H. and Acree, W. E. Estimation of enthalpies of sublimation of organic, organometallic and inorganic compounds. Fluid Phase Equilibria 2020, 515, 112575.
6. Heid, E. , Greenman, K. P. , Chung, Y. , Li, S. C. , Graff, D. E. , Vermeire, F. H. , Wu, H. , Green, W. H. , and McGill, C. J. Chemprop: A Machine Learning Package for Chemical Property Prediction. Journal of Chemical Information and Modeling 2024, 64, 9–17.
7. Yang, K. , Swanson, K. , Jin, W. , Coley, C. , Eiden, P. , Gao, H. , Guzman-Perez, A. , Hopper, T. , Kelley, B. , Mathea, M. , Palmer, A. , Settels, V. , Jaakkola, T. , Jensen, K. , and Barzilay, R. Analyzing Learned Molecular Representations for Property Prediction. Journal of Chemical Information and Modeling 2019, 59, 3370–3388.
8. Rittig, J. G. and Mitsos, A. Thermodynamics-consistent graph neural networks. Chemical Science 2024, 15, 18504–18512.
9. Specht, T. , Nagda, M. , Fellenz, S. , Mandt, S. , Hasse, H. , and Jirasek, F. HANNA: hard-constraint neural network for consistent activity coefficient prediction. Chemical Science 2024, 15, 19777–19786.
10. Acree, W. and Chickos, J. S. Phase transition enthalpy measurements of organic and organometallic compounds and ionic liquids. sublimation, vaporization, and fusion enthalpies from 1880 to 2015. part 2. C11-C192. Journal of Physical and Chemical Reference Data 2017, 46
11. Acree, W. and Chickos, J. S. Phase Transition Enthalpy Measurements of Organic and Organometallic Compounds. Sublimation, Vaporization and Fusion Enthalpies From 1880 to 2015. Part 1. C1-C10. Journal of Physical and Chemical Reference Data 2016, 45
12. Acree, W. and Chickos, J. S. Phase Transition Enthalpy Measurements of Organic Compounds. An Update of Sublimation, Vaporization, and Fusion Enthalpies from 2016 to 2021. Journal of Physical and Chemical Reference Data 2022, 51
13. Kim, S. , Chen, J. , Cheng, T. , Gindulyte, A. , He, J. , He, S. , Li, Q. , Shoemaker, B. A. , Thiessen, P. A. , Yu, B. , Zaslavsky, L. , Zhang, J. , and Bolton, E. E. PubChem 2025 update. Nucleic Acids Research 2025, 53, D1516–D1525.
14. NCI/CADD Group Chemoinformatics Tools and User Services. https://cactus.nci.nih.gov/ (accessed Jan. 24, 2025).
15. Chickos, J. S. , Hosseini, S. , Hesse, D. G. , and Liebman, J. F. Heat capacity corrections to a standard state: a comparison of new and some literature methods for organic liquids and solids. Structural Chemistry 1993, 4, 271–278.
16. Johnson, M. S. , Dong, X. , Grinberg Dana, A. , Chung, Y. , Farina, D. , Gillis, R. J. , Liu, M. , Yee, N. W. , Blondal, K. , Mazeau, E. , Grambow, C. A. , Payne, A. M. , Spiekermann, K. A. , Pang, H. W. , Goldsmith, C. F. , West, R. H. , and Green, W. H.

RMG Database for Chemical Property Prediction. Journal of Chemical Information and Modeling 2022, 62, 4906–4915.
17. Vermeire, F. H. and Green, W. H. Transfer learning for solvation free energies: From quantum chemistry to experiments. Chemical Engineering Journal 2021, 418, 129307.
18. Joback Group Contribution Method (thermo.group_contribution.joback) — Thermo 0.4.0 documentation. https://thermo.readthedocs.io/thermo.group_contribution.joback.html (accessed Jan. 22, 2025).
19. Lin, S. T. , Chang, J. , Wang, S. , Goddard, W. A. , and Sandler, S. I. Prediction of vapor pressures and enthalpies of vaporization using a COSMO solvation model. Journal of Physical Chemistry A 2004, 108, 7429–7439.
20. Chickos, J. S. , Nichols, G. , and Ruelle, P. The estimation of melting points and fusion enthalpies using experimental solubilities, estimated total phase change entropies, and mobile order and disorder theory. Journal of Chemical Information and Computer Sciences 2002, 42, 368–374.
21. Liu, Y. , Tran, H. , Huang, C. , Rio, B. G. del , Joseph, V. R. , Losego, M. , and Ramprasad, R. Accelerated predictions of the sublimation enthalpy of organic materials with machine learning. Materials Genome Engineering Advances 2025, e84.