Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

A Modeling Framework Integrating Data Trends and Reference Information for Predicting Temperature-Dependent Thermophysical Properties

Shuai Zhang[a], Abdulelah S. Alshehri[b*], Mansour S. Alhoshan[b], and Anjan Tula[a*]

[a] State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China.
[b] Chemical Engineering Department, College of Engineering, King Saud University, Riyadh 11421, KSA
* Corresponding author: mhoshan@KSU.EDU.SA.
* Corresponding author: anjantula@zju.edu.cn.

ABSTRACT

The availability of temperature-dependent physicochemical property data forms the cornerstone of process simulation, optimization, and sustainable molecular and product design. However, a critical data gap persists, as experimental measurements are accessible for only a small subset of known chemicals. This renders experimental characterization resource-prohibitive, often compelling reliance on empirical estimation methods. Moreover, although many models offer single-point predictions at fixed temperatures, accurately modeling continuous temperature-dependent behavior remains challenging. Conventional methods frequently overlook intermediate variations, resulting in limited extrapolation capability. To overcome these limitations, we introduce a mechanism-guided hybrid modeling framework that integrates physical insights into data-driven models. This framework is built on two strategies. Strategy I targets trend correction by generating a continuous representation from discrete single-point predictions, incorporating descriptors and slopes. Strategy Ⅱ addresses bias removal by anchoring a baseline to a high-accuracy point estimate and fitting the remaining deviations. The framework's effectiveness is evidenced by evaluations across ten thermophysical properties: Strategy I achieves MSE reductions of 19.23% and 20.33% for the quantitative structure-property relationship and group contribution methods, respectively. Strategy II provides a more substantial improvement, attaining an 81.63% MSE reduction for the gradient boosting decision tree regression model. This work demonstrates that incorporating trend and slope constraints facilitates physically consistent, bias-corrected, and accurate predictions, offering a scalable approach to bridge the data gap and accelerate computer-aided engineering and design.

Keywords: Temperature-dependent property prediction, Hybrid modeling, Mechanistic constraints, Machine learning, Slope-based correction, Bias correction

1. INTRODUCTION

The characteristics of organic compounds, especially the physical properties related to energy and phase behavior, are of great significance in the field of process engineering and serve as the foundation for process design and optimization. However, due to the scarcity of experimental data and the high cost of conducting experimental measurements, property prediction models thus represent the most viable and practical means to estimate the required properties in the initial design phase.

Benefiting from the relative abundance of standardized data points, single-point prediction models (such as boiling point and critical temperature) have seen significant development. In stark contrast, the development of models for temperature-dependent properties remains a notably under-explored area. This research gap is not only due to the scarcity of comprehensive experimental

data but, more fundamentally, to the significant increase in modeling complexity. Predicting a single value, by comparison, constitutes a fundamentally simpler task than learning a continuous temperature-dependent function. The latter demands models that can simultaneously capture the interplay of molecular structure and thermal dynamics, adhere to underlying physical laws, and maintain generalizability across both chemical space and continuous temperature domains. This pronounced complexity, coupled with the scarcity of comprehensive data, has therefore intensified the need for accurate and efficient models capable of estimating thermophysical properties for compounds lacking experimental measurements.

The most classic predictive models include: group contribution (GC) models[1], quantitative structure-property relationship (QSPR) models[2], and machine learning (ML)models[3]. In the prediction of temperature-dependent properties, GC models are often used to regress the coefficients of thermodynamic approximation or correlation functions. This workflow and its essentials are well captured by early foundational work [4]. Early temperature-dependent predictive models include:

(1)Heat capacity polynomial

$$C_p = [\sum(a) - 37.93] + [\sum(b) + 0.210]T + [\sum(c) - 3.91 \times 10^{-4}]T^2 + [\sum(d) + 2.06 \times 10^{-7}]T^3 \quad (1)$$

(2)Liquid-viscosity (Andrade-type) relation

$$\eta_L = mw \times \exp\left\{\frac{\sum(\eta_a) - 597.82}{T} + \sum(\eta_b) - 11.202\right\} \quad (2)$$

For models without explicit functional forms, a standard approach to predicting temperature-dependent properties is to treat temperature as an additional input feature, concatenated with molecular descriptors or structural information. This approach is widely used in various modeling contexts. For instance, in QSPR modeling, Sosnowska et al. [5] integrated QSPR-predicted enthalpy of vaporization into the Clausius-Clapeyron equation to characterize the temperature-dependent vapor pressure of persistent organic pollutants. This data-driven strategy has enabled predictions even under extreme conditions, as demonstrated by Yin et al. [6], who predicted gas heat capacity across an exceptionally wide temperature range. Similarly, tree-based ensemble methods such as gradient boosting regression trees and light gradient boosting machine have been employed to predict liquid heat capacity [7] and solubility [8] across varying temperatures by treating temperature as an input variable.

These models share a fundamental limitation in handling temperature-dependent properties: the prevalent methodology essentially treats data points from the same compound at different temperatures merely as an expanded dataset, processing each (temperature, property) pair as an independent sample. This approach inherently disregards the intrinsic physical continuity of the temperature-property curve. Consequently, conventional models trained on such discrete (structure, temperature) pairs fail to learn the underlying continuous functional relationship, resulting in physically inconsistent predictions that often exhibit incorrect curve shapes and poor extrapolation beyond the training temperature domain.

The above theoretical limitations are clearly reflected in practical predictions. As shown in Figure 1, the GC model, constrained by its fixed functional form, fails to accurately capture experimental trends: Figure 1A demonstrates its constant-trend prediction for heat capacity that deviates from the experimental data, while Figure 1B and 1C show increasingly amplified errors as the temperature range extends. Figure 1D further illustrates its overall failure to follow the actual property trend.

In contrast, while QSPR models offer greater flexibility, they introduce a different challenge due to their purely data-driven nature. As shown in Figure 2, all four subfigures (A-D) display abrupt, physically unsupported changes in predictions. Notably, Figures 2B and 2C exhibit nearly non-monotonic segments that contradict fundamental thermodynamic principles.

2. PROPOSED HYBRID MODELING FRAMEWORK

2.1 Strategy I – Trend-Guided modeling

Table 1: Linearity assessment of ten temperature dependent thermophysical properties. Data were obtained from the pure component analysis module of Aspen Plus. For each property, the R^2 value is averaged across all substances, where each substance's R^2 is calculated from the linear regression of property values over temperature points.

Property (abbreviation)	Mean R^2	Count Of R^2<0.95
Heat capacity(C_p)	0.9768	27*
Heat of vaporization($\Delta Hvap$)	0.9843	8
Internal energy(U)	0.9932	0
Liquid density(ρ_L)	0.9904	0
Pure component enthalpy(H)	0.9948	0
Pure component exergy(Ex)	0.9909	0
Gibbs free energy(G)	0.9964	0
Isentropic exponent(γ)	0.9927	0
Vapor pressure(P_{vap})	0.9942	0
Volume(V)	0.9731	4

*17 substances with R^2 between 0.9 and 0.95, 6 substances with R^2 between 0.8 and 0.9, and 4

substances with R^2 below 0.8.

As demonstrated in the table above, fitting a linear trend to the experimental data for ten temperature-dependent properties reveals a strong linear relationship in all cases, with an average coefficient of determination (R^2) exceeding 0.97. (Note that vapor pressure data was log-transformed prior to fitting.) Consequently, high-accuracy predictions at two physically meaningful temperature points, such as the critical temperature (T_c) and the normal boiling point (T_b), are employed to estimate the property's temperature-dependent tendency.

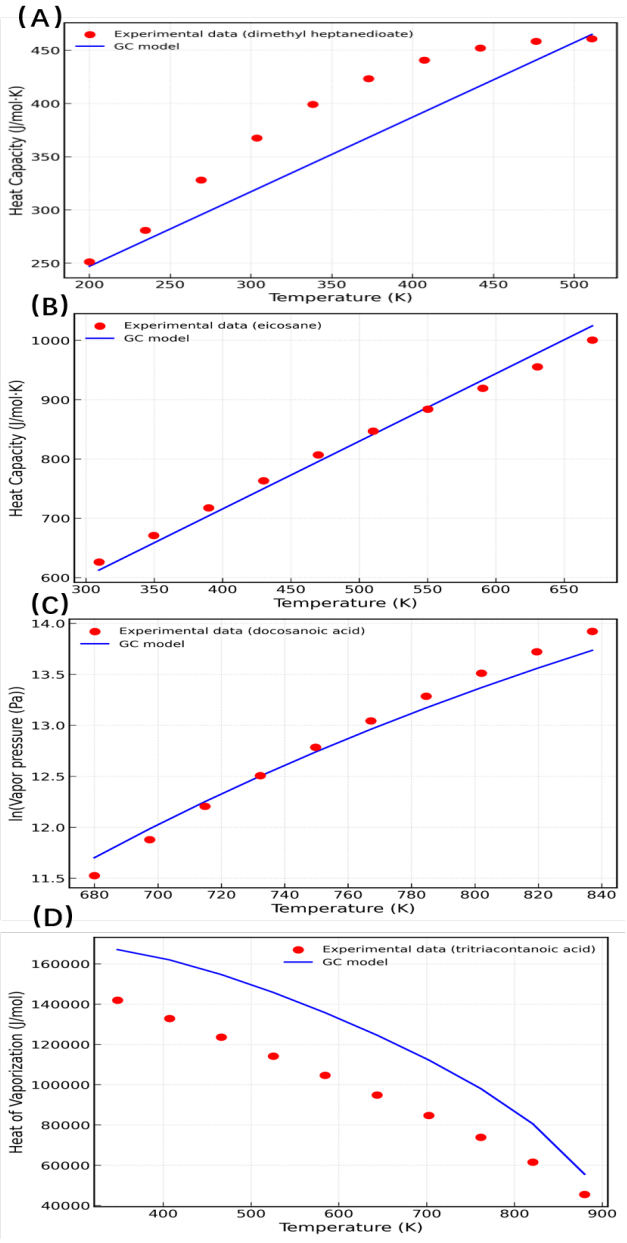

Figure 1. Prediction Performance of the GC Model for Thermodynamic Properties.(A, B) Heat capacity[9]. (C)Vapor pressure[10].(D)Heat of vaporization[9]. Model details are available in the Supplementary Information.

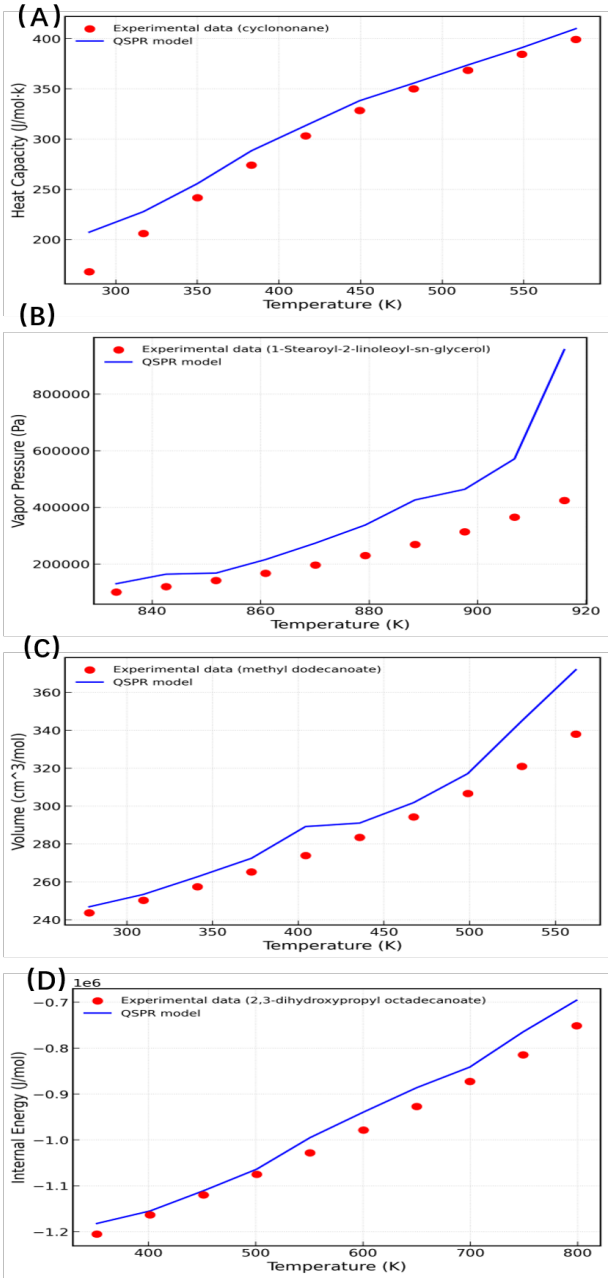

Figure 2. Prediction Performance of the QSPR Model for Thermodynamic Properties.(A) Heat capacity.(B)Vapor pressure.(C) Volume. (D) Internal Energy. Molecular descriptors derived from 2D structures downloaded from PubChem were used as input features. After data cleaning, appropriate descriptors were selected, and a random forest regression model was employed. Model details are available in the Supplementary Information.

2.1.1 Application of strategy I for explicit models

As presented in Figure 3, the modeling of temperature dependence for explicit models is accomplished through the following two-step framework:

Step 1: Construct four foundational models from molecular structure: two for predicting the reference

temperatures T_1 and T_2, and two for estimating the corresponding properties y_{r1} and y_{r2} at these points. The slope between these reference points is then derived as follows:

$$tendency = \frac{y_{r2}-y_{r1}}{T_2-T_1} \quad (3)$$

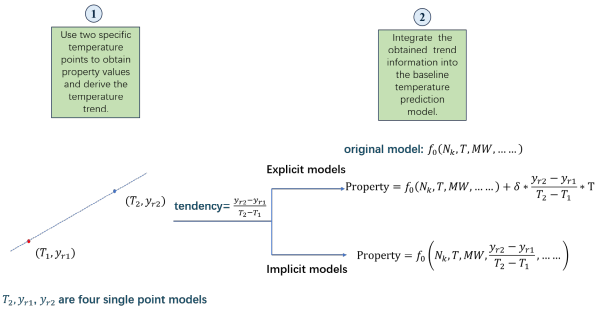

Figure 3. Flowchart of the Trend-Guided Hybrid Modeling Framework. Step 1: Construct models to predict two reference temperature points and their corresponding target properties, then use these four models to derive the predicted slope. Step 2: For existing models with an explicit expression, incorporate this slope as an additional temperature interaction term. For models without an explicit expression, use the slope as an additional input feature.

Step 2: Incorporate this slope into a baseline explicit model by introducing a temperature correction term. The augmented model takes the form:

$$\hat{y} = f_0(N_k, T, MW, \dots\dots) + \delta * \frac{y_{r2}-y_{r1}}{T_2-T_1} * \mathrm{T} \quad (4)$$

Here, $f_0(N_k, T, MW, \dots\dots)$denotes the original mechanistic or empirical model. Its temperature-inclined correction is introduced via a term weighted by a coefficient δ and scaled by the temperature T. This design allows the model to adaptively capture temperature-driven behavior.

2.1.2 Application of strategy I for ML models

The temperature-dependent modeling for ML models follows a two-step procedure:

Step 1: Construct four foundational models from molecular structure: two for predicting the reference temperatures T_1 and T_2, and two for estimating the corresponding properties y_{r1} and y_{r2} at these points. The slope between these reference points is then derived as follows:

$$tendency = \frac{y_{r2}-y_{r1}}{T_2-T_1} \quad (5)$$

Step 2: Incorporate the temperature tendency by adding the slope as a new input feature to the original data-driven model. This allows the model to learn temperature-dependent relationships directly, while its architecture remains unmodified. The resulting enhanced model formulation is:

$$\hat{y} = f_0\left(N_k, T, MW, \dots\dots, \frac{y_{r2}-y_{r1}}{T_2-T_1}\right) \quad (6)$$

Here, $f_0(N_k, T, MW, \dots\dots)$represents the original implicit model that learns structure-property relationships directly from data.

2.1.3 Physical significance

The physical significance of Strategy I stems from its fundamental departure from discrete-point prediction. By introducing a "temperature tendency" derived directly from experimental data points, the method incorporates prior physical intuition about continuous, smooth property evolution. This data-driven yet physics-aware approach not only improves robustness against experimental noise but also ensures more reliable extrapolation and interpolation. Crucially, it bridges the gap between discrete experimental observations and continuous theoretical physical models.

2.2 Strategy II –Bias-Corrected hybrid modeling

The conceptual foundation of Strategy II originates from a critical observation regarding purely data-driven machine learning methods. While these algorithms excel at identifying complex patterns, we propose that a more effective approach is to first establish a predictive baseline model, rather than feeding all data directly into a black-box model. This baseline, which uses linear regression with group features as inputs, captures the fundamental trend; subsequently, a GBDT model is applied specifically to learn the deviations between the baseline predictions and the actual measured values. This two-step process is illustrated in the flowchart of Figure 4.

2.2.1 Baseline model construction

Supported by the linearity assessment of ten temperature-dependent thermophysical properties (Table 1), a linear model is adopted as the baseline. Two preliminary models are first built to predict a reference temperature, T_1, and the corresponding property value at that temperature, y_{r1}, directly from molecular structure. This reference prediction, y_{r1}, serves as the intercept of the linear baseline model. The slope of the baseline is then determined using group contribution methods, giving the following expression for the baseline prediction $\hat{y}$:

$$\hat{y} = y_{\mathrm{r1}} + \sum_K A_K N_K (T - T_1) \quad (7)$$

Here, y_{r1} is the reference prediction for the property at temperature T_1, A_K denotes the group contribution coefficients, N_K represents the counts of molecular groups, and $T - T_1$ is the temperature deviation from the reference point.

2.2.2 Bias learning and compensation

Although the baseline model is physically reasonable, it cannot fully capture all complex, nonlinear relationships. To address this limitation, the deviation between the baseline prediction and the actual experimental value is calculated as $b = y - \hat{y}$. A machine learning model is subsequently employed to predict this systematic bias:

$$\hat{b} = f_1(N_k, T, MW \dots \dots) \tag{8}$$

This model, f_1, learns the residual errors of the baseline using a set of features including molecular descriptors, temperature, and molecular weight. The final, refined prediction is obtained by compensating the baseline output with the predicted bias:

$$\check{y} = \hat{y} + b \tag{9}$$

This step combines the robustness of the physical baseline with the flexibility of machine learning, enabling the hybrid model to capture nonlinear deviations that the baseline alone cannot represent.

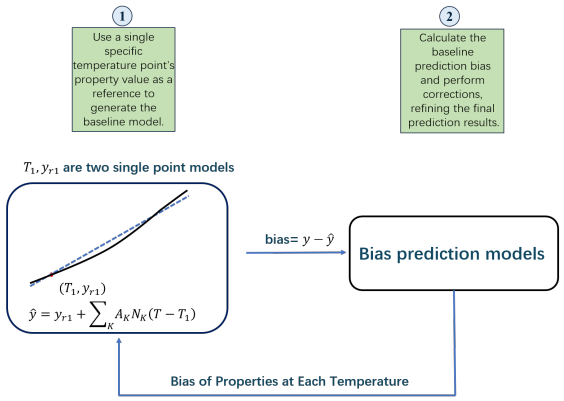

Figure 4. Flowchart of the Bias-Corrected Hybrid Modeling Framework. Step 1: Construct a prediction model for a reference temperature point and its corresponding target property to serve as an anchor point. Then, using the group contribution method, regress a straight line through this anchor point, where the method determines the slope. This line is taken as the predictive prior model for the property. Step 2: Calculate the deviation between the predictions of this prior model and the actual observed values. By predicting this deviation, apply a feedback compensation to the baseline model.

2.2.3 Physical significance

This hybrid methodology offers distinct advantages over purely black-box machine learning. Its primary strength lies in enforcing physically realistic predictions by anchoring them to a group-contribution baseline that adheres to thermodynamic principles. This constraint mitigates the generation of unrealistic outcomes—a common issue in purely data-driven models, particularly during extrapolation beyond the training domain. The framework also enhances interpretability: the baseline provides a physically meaningful starting point, while the machine learning component is dedicated to modeling only the residual deviations from this baseline. This separation allows researchers to distinguish between the fundamental physical relationships embedded in the baseline and the more complex, data-driven patterns captured by the machine learning model. Furthermore, by decomposing the problem into these two complementary tasks, the approach achieves greater data efficiency and robustness, as the machine learning model is tasked specifically with correcting the baseline's limitations rather than learning the entire underlying relationship from scratch.

3. RESULTS AND DISCUSSION

3.1 Data and evaluation framework

Table 2: Summary of experimental data points for the ten evaluated temperature-dependent properties, For each substance, data was collected at 10 different temperature points. Experimental data were obtained from Aspen Plus.

Property	Symbol	Datapoints
C_p	$J/mol \cdot k$	2070
$\Delta Hvap$	J/mol	2040
U	J/mol	2070
ρ_L	g/ml	2070
H	J/mol	2070
Ex	J/mol	2050
G	J/mol	2050
γ	1	2070
P_{vap}	Pa	2090
V	cm^3/mol	2080

To evaluate the proposed enhancement strategies, we conducted a systematic performance comparison across several modeling frameworks. In this notation, the suffix "_SI" indicates the application of Strategy I, and "_SII" indicates Strategy II. Using this labeling scheme, we compared the following model variants: GC vs. GC_SI, GBDT vs. GBDT_SII, and QSPR vs. QSPR_ SI. All model variants are clearly identified with these labels in the subsequent results and discussion sections.

Table 2 summarizes the experimental data points for the ten evaluated temperature-dependent properties. For each substance, data were collected at ten distinct temperature points to comprehensively capture the property's behavior across a temperature range. These experimental data, serving as the foundation for all evaluations, were sourced from Aspen Plus—a widely used process simulation tool known for providing reliable thermophysical property data for diverse compounds.

Table 3. Overall predictive performance of original and enhanced models for ten thermophysical properties. Original models are labeled as GC, GBDT, and QSPR; models enhanced by Strategy I and Strategy II are denoted by the suffixes "_SI" and "_SII", respectively. For specific models and raw data, please refer to the digital supplementary material.

Properties		GC	GC_SI	GBDT	GBDT_SII	QSPR	QSPR_SI
C_p	R^2	0.9918	0.9924	0.9971	0.9981	0.9994	0.9994
	MSE	16.08	14.89	5.60	3.66	1.19	1.14
P_{vap}	R^2	0.9033	0.9194	0.8476	0.9914	0.9886	0.9901
	MSE	63.04	52.58	99.36	5.59	7.41	6.43
$\Delta Hvap$	R^2	0.9260	0.9534	0.9713	0.9941	0.9929	0.9961
	MSE	64.55	40.66	24.98	5.14	6.15	3.40
V	R^2			0.9931	0.9985	0.9991	0.9992
	MSE			331.45	73.67	41.83	37.78
U	R^2			0.9965	0.9996	0.9991	0.9992
	MSE			50.58	6.36	12.59	11.51
H	R^2			0.9965	0.9996	0.9991	0.9992
	MSE			49.85	6.18	13.17	11.00
Ex	R^2			0.9728	0.9987	0.9992	0.9992
	MSE			119.77	5.90	3.59	3.53
γ	R^2			0.9600	0.9956	0.9931	0.9955
	MSE			354.26	39.37	61.24	40.08
ρ_L	R^2			0.9636	0.9957	0.9932	0.9950
	MSE			67.49	8.00	12.53	9.23
G	R^2			0.9843	0.9975	0.9957	0.9971
	MSE			50.27	8.05	13.80	9.27

3.2 Comparative performance analysis

Model performance was quantitatively assessed using the R^2 to evaluate the overall goodness of fit and the MSE to measure the magnitude of absolute error.

The overall performance of the models in predicting the ten target properties is summarized in Table 3. Note that results from the GC-based model are only reported for heat capacity, vapor pressure, and heat of vaporization; for the remaining properties, no established GC models were available, and those entries are left blank. Based on the results in Table 3, a consistent improvement is observed across all original models after incorporating the proposed strategies, evidenced by increased R^2 values and reduced MSE. This dual improvement reflects both a better overall fit and a systematic reduction in prediction errors, thereby statistically validating the effectiveness of Strategy I and Strategy II.

Table 4: MSE improvement results using Strategy I and Strategy II: Percentage improvement in MSE relative to the original model for each property. Blank entries indicate cases where the corresponding strategy-model combination was not applicable. Original models are denoted as GC, gradient boosting decision tree regression (GBDT), and QSPR; models enhanced with Strategy I and Strategy II are labeled with the suffixes "_SI" and "_SII", respectively.

Properties	GC_SI	GBDT_SII	QSPR_SI
C_p	7.40%	34.64%	4.20%
P_{vap}	16.59%	94.37%	13.23%
$\Delta Hvap$	37.01%	79.42%	44.72%
V		77.77%	9.68%
U		87.42%	8.58%
H		87.60%	16.48%
Ex		95.07%	1.67%
γ		88.89%	34.55%
ρ_L		88.15%	26.34%
G		83.99%	32.83%
Mean	20.33%	81.63%	19.23%

To further quantify the improvement, we compared the percentage reduction in MSE before and after applying the proposed strategies. As shown in Table 4, the strategies yield substantial and broadly consistent error reduction across most models. With Strategy I, the average MSE decreased by 19.23% for QSPR-based models and by 20.33% for GC models. For GBDT models, Strategy II delivered the most pronounced improvement, achieving an 81.63% reduction in MSE. These results underscore the efficacy of the Bias-Corrected Hybrid Modeling Framework for complex data-driven modeling tasks.

3.3 Model interpretation and physical insights

Figure 5 compares the original and enhanced models across three temperature-dependent properties. In Figure 5A, the GC model—originally constrained by a fixed linear form—exhibits markedly improved alignment with the experimental trend after incorporating Strategy I, yielding more flexible predictions that reduce deviations at both ends of the temperature range. Figure 5B shows that for the QSPR model, Strategy I preserves its inherent capacity to capture nonlinearity while imposing a physically guided constraint that suppresses unphysical, abrupt variations. The remaining panels consistently illustrate the advantages gained by integrating the proposed physics-informed features, further confirming the general effectiveness of the enhancement strategies.

The comparative results highlight a consistent advantage of incorporating physical knowledge into data-driven models. While purely statistical approaches learn patterns directly from the data, they lack awareness of the underlying thermodynamic structure governing temperature-dependent properties. The proposed strategies address this gap by embedding mechanistic priors into the learning process, enabling the hybrid framework to recover physically meaningful trends that conventional models often fail to capture. As illustrated by the predictions, many original models struggle to reproduce the experimental temperature variation because they are trained across multiple compounds using a global loss function that produces parameter sets representing statistical compromises. This leads to persistent prediction gaps for specific compounds—systematic biases that cannot be eliminated without incorporating physical mechanisms.

The physics-informed strategies introduced in this work effectively resolve these issues by embedding fundamental temperature-dependent behavior into the predictive process. In particular, Strategy I incorporates a physically interpretable slope term extracted from high-accuracy reference points such as T_b and T_c. This slope simultaneously acts as a physics-informed feature—encoding the direction and magnitude of temperature effects—and as a molecularly grounded descriptor, since it reflects influences such as intermolecular forces, molecular weight, functional group contributions, and polarizability. By doing so, Strategy I transforms the learning task from predicting an unconstrained surface in (structure, T) → property space to predicting deviations around a physically anchored linearized manifold, analogous to embedding a first-order thermodynamic Taylor expansion as a prior.

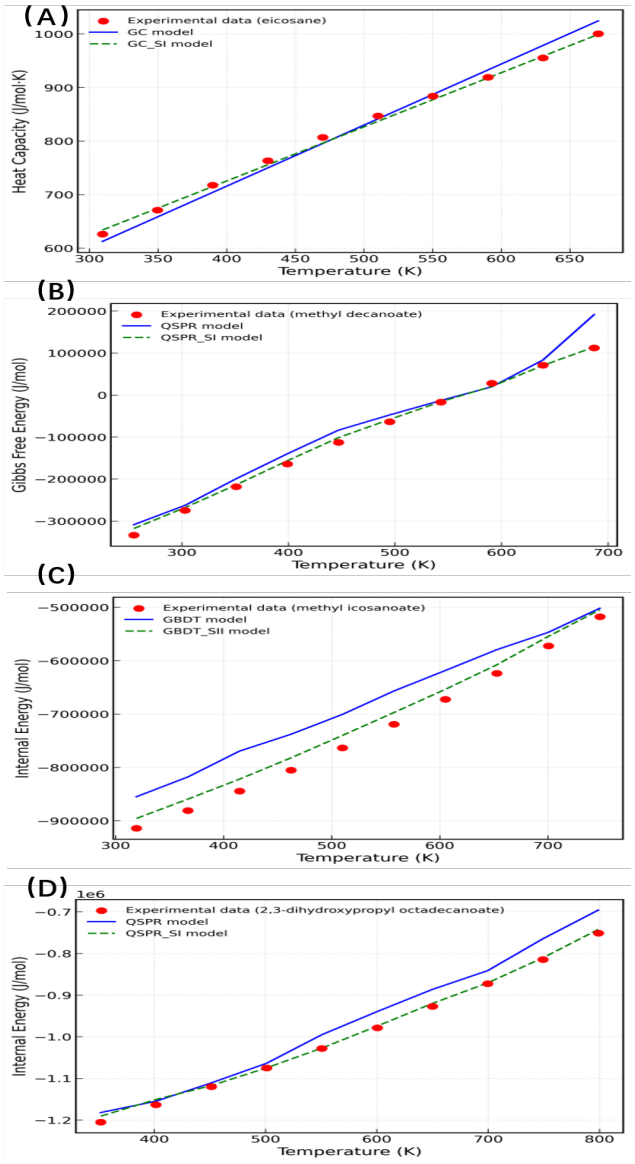

Figure 5. (A) GC vs. GC_SI for heat capacity. (B) QSPR vs. QSPR_SI for internal energy. (C) GBDT vs. GBDT_SII for internal energy. (D) QSPR vs. QSPR_SI for Gibbs free energy. (The suffix "_SI" denotes enhancement by Strategy I, "_SII" by Strategy II). For specific models and raw data, please refer to the digital supplementary material.

Similarly, Strategy II leverages a baseline GC model that encodes known physical relationships between

molecular groups and temperature tendencies. Although the baseline is approximate, it captures the dominant trend with clear physical interpretability, while the machine-learning component predicts only the deviation from this baseline. This decomposition substantially reduces the learning difficulty: the ML model focuses solely on complex residual patterns rather than the entire thermodynamic mapping. The baseline also prevents thermodynamic inconsistencies, such as incorrect concavity or unphysical changes in slope. As a result, the overall error decreases because ML is used to refine the nonlinear residual structure rather than reconstruct the entire temperature dependence from scratch. This approach closely mirrors the "Δ-learning" philosophy used in quantum chemistry machine learning, and our framework adopts the same logic for temperature-dependent thermophysical properties.

4 CONCLUSIONS

This paper presents a predictive framework for temperature-dependent properties that advances existing methodologies by addressing a fundamental limitation: the prevalent treatment of experimental data as discrete points rather than as manifestations of an underlying continuous function. By incorporating physically meaningful temperature trends and systematically correcting biases within data-driven workflows, the framework enables models to learn complete temperature-dependent functional relationships. Beyond predictive accuracy, it ensures thermodynamic consistency, enhances interpretability, and improves extrapolation reliability—attributes that are critical for process simulation, optimization, and computer-aided molecular design.

DIGITAL SUPPLEMENTARY MATERIAL

The complete source code, datasets, and implementation details for all models presented in this study have been made openly available. The repository contains the input features for all 10 properties across 200+ compounds, the code for all comparison models. The TRGPACK repository can be accessed at: https://github.com/SZ-ZJU/TRGPACK.

ACKNOWLEDGEMENTS

Financial support from the Natural Science Foundation of China (No. 22150410338) is gratefully acknowledged. The authors would like to extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through ISPP program (ISPP25-11).

REFERENCES

1. Hukkerikar AS, Sarup B, Ten Kate A, Abildskov J, Sin G, Gani R. Group-contribution+ (GC+) based estimation of properties of pure components: improved property estimation and uncertainty analysis. Fluid Phase Equilibria 321:25-43 (2012). https://doi.org/10.1016/j.fluid.2012.02.010
2. Yalamanchi KK, van Oudenhoven VCO, Tutino F, Monge-Palacios M, Alshehri A, Gao X, Sarathy SM. Machine learning to predict standard enthalpy of formation of hydrocarbons. J. Phys. Chem. A 123:8305-8313 (2019). https://doi.org/10.1021/acs.jpca.9b04771
3. Alshehri AS, Tula AK, You F, Gani R. Next generation pure component property estimation models: with and without machine learning techniques. AIChE Journal 68: (2021). https://doi.org/10.1002/aic.17469
4. JOBACK KG, REID RC. ESTIMATION OF PURE-COMPONENT PROPERTIES FROM GROUP-CONTRIBUTIONS. Chemical Engineering Communications 57:233-243 (2007). https://doi.org/10.1080/00986448708960487
5. Sosnowska A, Barycki M, Jagiello K, Haranczyk M, Gajewicz A, Kawai T, Suzuki N, Puzyn T. Predicting enthalpy of vaporization for persistent organic pollutants with quantitative structure–property relationship (QSPR) incorporating the influence of temperature on volatility. Atmospheric Environment 87:10-18 (2014). https://doi.org/10.1016/j.atmosenv.2013.12.036
6. Yin J, Jia Q, Yan F, Wang Q. Predicting heat capacity of gas for diverse organic compounds at different temperatures. Fluid Phase Equilibria 446:1-8 (2017). https://doi.org/10.1016/j.fluid.2017.05.006
7. Shan Y, Wu Q, Yuan H, Liu W. Develop machine learning-based model and automated process for predicting liquid heat capacity of organics at different temperatures. Fluid Phase Equilibria 584:114132 (2024). https://doi.org/10.1016/j.fluid.2024.114132
8. Ye Z, Ouyang D. Prediction of small-molecule compound solubility in organic solvents by machine learning algorithms. J Cheminform 13: (2021). https://doi.org/10.1186/s13321-021-00575-3
9. Ceriani R, Gani R, Meirelles AJA. Prediction of heat capacities and heats of vaporization of organic liquids by group contribution methods. Fluid Phase Equilibria 283:49-55 (2009). https://doi.org/10.1016/j.fluid.2009.05.016
10. Ceriani R, Meirelles AJA. Predicting vapor–liquid equilibria of fatty systems. Fluid Phase Equilibria 215:227-236 (2004).

https://doi.org/10.1016/j.fluid.2003.08.011