Developing a semi-empirical model for refrigerant compressors using symbolic regression: a data-driven approach for heat pump applications

Tom-Alexander Müller[a], Leon P. M. Brendel[b], Stefan S. Bertsch[b], Yixia Xu[a], Christiane Thomas[a]

[a]Technische Universität Dresden,
Schaufler Chair of Refrigeration, Cryogenics and Compressor Technology,
01062 Dresden, Germany
[b]OST – Eastern Switzerland University of Applied Sciences,
Institute for Energy Systems, 9741 Buch, Switzerland

Abstract

Accurate thermodynamic modeling of compressors is essential for the analysis, optimization and control of heat pump systems. Traditional modeling methods, such as the AHRI-540 10-coefficient polynomial, lack generalizability and are fluid-specific, requiring costly re-parameterization for different refrigerants. To address this limitation, this work explores symbolic regression (SR) as a transparent, data-driven technique to derive a refrigerant-generalized correlation for isentropic efficiency. This methodology was applied to a comprehensive experimental dataset from reciprocating compressors operating with a diverse range of fluids, including HFOs, HFCs, HCs and their mixtures. The SR search successfully discovered a novel, parsimonious correlation for isentropic efficiency that is a function of these physical properties. However, the current correlation still has limitations when it comes to correctly predicting performance over the entire compressor envelope, particularly in areas where there is little measuring data available. This paper presents the methodology, the proposed correlation and a discussion of its benefits and limitations. The physics-informed SR approach offers a promising path to generalizable models, though the proposed equation requires further refinement to achieve full topological accuracy across more refrigerant classes.

Keywords:Compressor modelling; System simulation; Symbolic regression;

1. Introduction

The decarbonization of the heating and cooling sector is a primary objective in global energy policy, with heat pumps representing a key enabling technology. The compressor is the component in the vapour-compression cycle that uses the most energy, and its performance dictates the overall efficiency of the system. Consequently, accurate system-level simulations, which are often conducted in platforms like TRNSYS or Modelica, are essential for component sizing, control strategy optimization, and reliable prediction of performance and system efficiency, such as seasonal coefficient of performance (SCOP). These system simulations requires component models, in particular compressor models that are accurate and require less calculation resource [1]. This study addresses the domain of fast-executing empirical or semi-empirical correlations, which are suitable for annual performance analysis. Computationally-intensive methods, such as detailed physical chamber models used for component design, are explicitly outside the scope of this work. Correlation based compressor models that are currently often implemented in system simulations are the 10-coefficient polynomial (AHRI-540), and it possesses significant limitations. The 10-coefficient polynomial is an empirical black box fit, is notoriously unreliable for extrapolation beyond its training data, and is inherently fluid-specific [2]. A performance map for one refrigerant operating at a fixed motor speed and a constant superheat, such as R1234yf (an HFO), is completely invalid for another, like R290 (an HC), requiring a costly and time-consuming re-parameterization for every new working fluid. This gap in modeling capability leads to a central research question: Is it possible to develop a single, refrigerant-generalized correlation that can accurately predict compressor performance for a wide range of working fluids, including pure fluids and mixtures, by incorporating fundamental thermophysical properties?

The primary limitation of existing correlations is their limitation to not explicitly incorporate fundamental properties of the working fluid, relying solely on operating conditions (e.g., temperatures, pressures). To overcome this, we propose a new methodology based on physics-informed symbolic regression (SR). SR is a machine learning paradigm that seeks to discover the underlying mathematical expression from data, balancing both accuracy and simplicity. We employ PySR [3], a high-performance, open-source SR library, to search for a new correlation.

The contribution of this work is the development and validation of a novel, refrigerant-generalized compressor efficiency correlation derived from this methodology. Using an experimental dataset covering reciprocating compressors and a diverse range of fluids (HFOs, HFCs, HCs and their mixtures), we present a new, parsimonious equation for isentropic efficiency. We then validate this new correlation, comparing its accuracy against Doughty et al. correlation and the AHRI-540 benchmark, and conclude by analyzing the physical interpretability and current limitations of the proposed equation.

2. Methodology

This section details the methodology, specifically by describing the symbolic regression framework PySR, the input variables for the framework and the experimental dataset.

2.1 Symbolic Regression with PySR

In this study, symbolic regression (SR) is employed to derive an equation capable of describing refrigerant-dependent behavior based on thermophysical properties. The selection of SR is driven by its ability to derive mathematical expressions optimized to balance predictive accuracy against model complexity for a specific dataset. This capability is crucial for scientific inquiry, which, diverging from many machine learning applications, places a stringent requirement on model explainability and verifiability. SR fulfills this need by generating human-interpretable formulations, offering a transparent alternative to opaque black box architectures (e.g., deep neural networks for predicting the pressure drop in microchannels [4]).

PySR [3] was chosen to find a mathematical function. PySR uses a genetic programming approach with an evolutionary algorithm. In each generation, new expressions are created through genetic operations such as mutation and crossover. These are then algebraically simplified, and their constants are numerically optimized. Therefore, with this approach a vast majority of possible correlations can be explored and evaluated with the dataset. A tree structure in symbolic regression graphically represents a mathematical expression. This hierarchical representation utilizes internal nodes for mathematical operators (e.g., +, ×, log), and terminal nodes (leaves) for variables and numerical constants. Crucially, the tree structure enables evolutionary algorithms, such as used in PySR, to efficiently manipulate, combine, and evolve candidate formulas, thus providing an unambiguous and executable format for the discovered correlation. In this study, the set of allowed mathematical operators was constrained. The set included the operators addition (+), subtraction (-), multiplication (*), division (/), and power (^). To guide the evolutionary search, the objective function was defined as the Mean Squared Error (MSE).

2.2 Feature Selection for Correlation Development

The selection of input variables is a critical step in developing a physically meaningful and robust correlation. To ensure the discovered correlation is based on fundamental principles, the input features were chosen to be unique and unambiguous properties of the refrigerant state.

All thermodynamic properties for the given experimental conditions were calculated using the open-source library CoolProp [5] for pure fluids and RefProp [6] for mixtures. The reason for this is that the correlation must be based on properties that uniquely define the thermodynamic state of a refrigerant or refrigerant mixture. By using refrigerant dependent properties (such as critical temperature and critical pressure, or compressibility factor and dynamic viscosity), we provide the symbolic regression algorithm with a physical basis. This approach reduces ambiguity and yields correlations that are physically interpretable rather than purely empirical. For every state point thermodynamic state and transport properties were calculated or measured. The properties are listed in Table 1.

Table 1: Considered thermodynamic influences as factors for correlation development

Symbol	Description	Unit	Function of
Calculated			
T_{crit}	Critical temperature	K	Refrigerant
p_{crit}	Critical pressure	Pa	Refrigerant
ρ_{crit}	Density at the critical point	kg m^{-3}	Refrigerant
μ_{suc}	Dynamic viscosity at suction state	Pa·s	T_{suc}, P_{suc}, Refrigerant
Z_{suc}	Compressibility factor at suction state	-	T_{suc}, P_{suc}, Refrigerant
ω	Acentric factor	-	Refrigerant
T_0	Evaporation temperature (dew point)	K	P_{suc}, Refrigerant
T_c	Condensation temperature (bubble point)	K	P_{suc}, Refrigerant
k_{suc}	Isentropic exponent at suction state	-	T_{suc}, P_{suc}, Refrigerant
P_{ratio}	Pressure ratio	-	P_{dis}/P_{suc}
ΔT_{sh}	Superheat at suction	K	$T_{suc} - T_0$
Measured			
P_{dis}	Discharge pressure	Pa	Refrigerant
P_{suc}	Suction pressure	Pa	Refrigerant
T_{suc}	Suction temperature	K	Refrigerant

2.2 Experimental Dataset

The data for this study were obtained from measurements of [7,8]. The experimental heat pump test bench is described in [7]. Two reciprocating compressors possessing similar swept volumes (0.158 l and 0.153 l) and similar outer shape and geometry were used. Tests involving synthetic refrigerants were conducted using the polyester oil, Reniso Triton SE 170, whereas the tests utilizing hydrocarbon refrigerants employed the polyalkylene glycol oil, Reniso LPG 150. It is important to notice, that hydrocarbons refrigerants were used only in one compressor whereas synthetic refrigerants were tested on both compressors. Figure 1 shows the distribution of the experimental dataset across the operating range of evaporation pressure and condensation pressure categorized by refrigerant class (synthetic refrigerants and hydrocarbons).

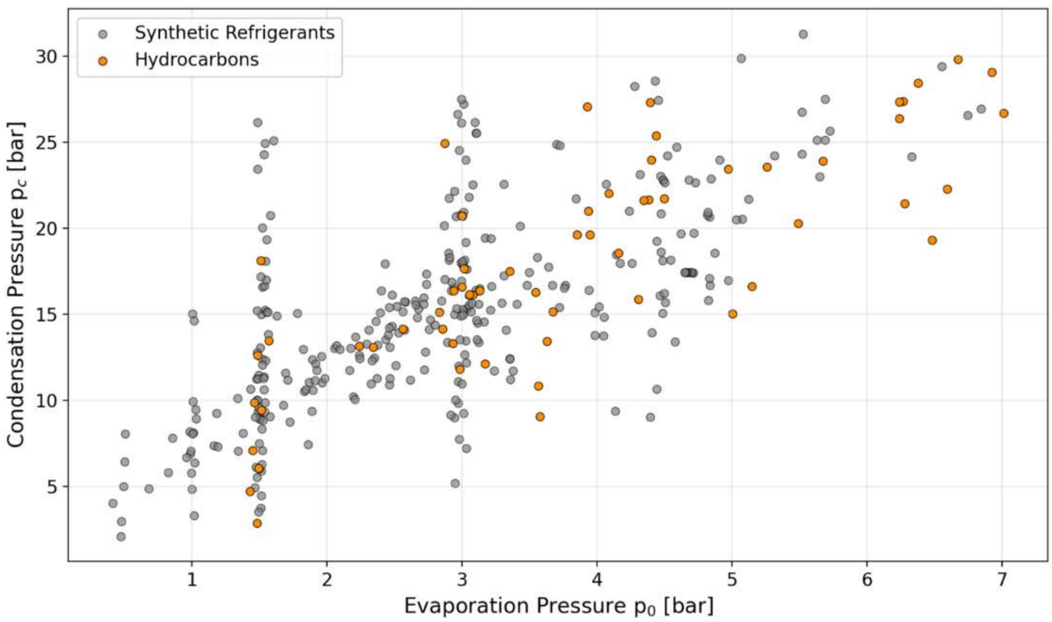

Figure 1: Distribution of the experimental dataset categorized by refrigerant class

A correlation derived by O. Doughty et. al from the dataset is presented in Equation 1. The primary advantage is that this equation offers good coverage across the operating range of the datasets.

$$\eta_{ois} = a_0 - \frac{a_1}{\pi^{a_4 \cdot P_{suc}}} - a_2 \cdot \pi^{R_x} + a_3 \cdot \Delta T_{sh} \tag{1}$$

Table 2: Coefficients for correlation from [8]

Coefficients	a_0	a_1	a_2	a_3	a_4	R_{syn}	R_{HC}
Value	0.6462	0.5798	0.0012	0.0012	0.0077	1.712	2.047

Table 2 shows the coefficients for Equation 1. To allow the correlation to uniquely fit the dataset for to synthetic and hydrocarbon refrigerants a refrigerant-dependent coefficient R_x is introduced. However, the value of R_x and also its role in Equation 1 is purely empirical and not guided by physical reasoning.

All refrigerants and mixtures used in the dataset from [8] are listed in Table 3 with the respective refrigerant class and calculated acentric factor ω and compressibility factor Z_{suc} at suction state for evaluating Equation 3. The acentric factor ω is a dimensionless number that measures the non-sphericity or structural complexity of a molecule. It quantifies the deviation of a fluid's thermodynamic properties, particularly its vapor pressure curve [9]. The compressibility factor Z_{suc} is a dimensionless quantity that quantifies the deviation of a real gas from ideal gas behavior. If $Z_{suc} > 1$, the gas is less compressible than an ideal gas due to repulsive forces, its volume is larger than expected. If $Z_{suc} < 1$, the gas is more compressible due to attractive forces, its volume is smaller than expected [10]. The use of these two dimensionless factors in Equation 3 is could be interpreted as the integration of real gas behaviour. By explicitly accounting for molecular non-sphericity through the acentric factor (ω) and intermolecular force effects through the compressibility factor (Z). To calculate the acentric factor of mixtures equation 5-3.3 from [3] was used and is hereby shown in Equation 2. Here, y_i represents the mole fraction of component i, and ω_i is its corresponding acentric factor. All thermophysical properties were calculated with CoolProp 7.0. [5] and the REFPROP Backend with Version 10.0 [6].

$$\omega_m = \sum_{i=1}^{n} y_i \cdot \omega_i \tag{2}$$

Table 3: Refrigerants from datasets in [7,8] with their respective $\pi, \omega, Z_{suc}, \Delta T_{sh}$

Refrigerant [Class]	$Pressure\ Ratio$ π $[-]$	Acentric factor ω $[-]$	$Compressibility$ $factor\ Z_{suc}$ $[-]$	Superheat ΔT_{sh} $[K]$
R1224yd(Z) [HFO]	3.14 - 8.67	0.322	0.911 - 0.970	7.72 - 52.08
R1234yf [HFO]	3.40 - 15.64	0.276	0.895 - 0.959	4.89 - 50.75
R1233zd(E) [HFO]	1.76 - 12.45	0.3025	0.896 - 0.980	13.01 - 46.45
R1336mzz(Z) [HFO]	2.36 - 9.73	0.386	0.912 - 0.982	15.96 - 46.68
R134a [HFC]	2.26 - 5.55	0.3268	0.906 - 0.943	16.16 - 49.98
R600 (Butane) [HC]	1.95 - 12.00	0.201	0.914 - 0.958	15.11 - 53.48
Binary mixtures				
R32 R1224yd(Z) [HFC HFO]	5.55 - 17.59	0.289 - 0.321	0.922 - 0.968	14.02 - 48.52
R1234yf R1233zd(E) [HFO / HFO]	2.38 - 16.18	0.291 - 0.294	0.904 - 0.982	11.74 - 44.18
R1234yf R1336mzz(Z) [HFO / HFO]	3.36 - 15.86	0.298 - 0.365	0.896 - 0.975	4.97 - 38.89
R1234yf R1224yd(Z) [HFO / HFO]	3.28 - 8.23	0.304	0.907 - 0.952	16.43 - 46.58
R290 R600 [HC / HC]	2.98 - 8.50	0.172	0.897 - 0.960	16.28 - 48.69
R290 R601 [HC / HC]	4.98 - 6.89	0.180	0.909 - 0.947	15.53 - 34.84

Ternary mixtures	$Pressure\ Ratio$ π [−}	Acentric factor ω [−]	$Compressibility$ $factor\ Z_{suc}$ [−]	Superheat ΔT_{sh} $[K]$
R32 R1224yd(Z) R1336mzz(Z) [HFC / HFO / HFO]	3.40 - 9.73	0.316	0.929 - 0.965	27.68 - 31.59
R32 R1234yf R1224yd(Z) [HFC / HFO / HFO]	4.31 - 6.89	0.292 - 0.309	0.899 - 0.958	14.80 - 39.74
R32 R1234yf R1336mzz(Z) [HFC / HFO / HFO]	4.85 - 7.11	0.277	0.906 - 0.955	25.65 - 32.46
R1234yf R1224yd(Z) R1336mzz(Z) [HFO / HFO / HFO]	3.40 - 9.73	0.328	0.922 - 0.969	16.30 - 45.13

3. Results and Discussion

In this section, we discuss the proposed compressor correlation for the isentropic efficiency, analyze its structure, and validate its accuracy against other correlations.

3.1 The proposed compressor correlation

Although the algorithm derived several regression equations, this paper focuses on one chosen correlation (Equation 3), which was selected for offering the optimal trade-off between accuracy and simplicity**.** The coefficients $a_0 = 0.595$ and $a_1 = 61.4$ were fitted with the dataset from [8].

$$\eta_{ois} = a_0 + \frac{\Delta T_{sh}}{T_{crit}} - {Z_{suc}}^{a_1} - \omega^{\pi} \tag{3}$$

The correlation is a sum of four terms. The first term is the linear offset, the coefficient a_0 could be fitted to describe deviations among various compressor models. The second term defines the normalized superheat by relating the actual superheat temperature difference to the refrigerant’s absolute critical temperature. It serves as a dimensionless parameter quantifying how far the gas is from its saturation point relative to its critical point. The third term subtracts the compressibility factor Z with an exponent a_1. This means a deviation from the ideal gas behavior at the compressor inlet. The last term is the acentric factor to the power of the pressure ratio π.

3.2 Model Validation

In this section, a systematic comparison is conducted for all refrigerants listed in Table 3. The analysis progresses from a qualitative visual inspection of the contour plots to a quantitative error metric, culminating in a physical interpretation of the observed differences. The mean average error and maximum absolute error are shown in Table 4. A comparison of the Mean Absolute Error (MAE) indicates that the Doughty et al. correlation demonstrates a nominally lower error compared to Equation 3. However, it is critical to say that these error magnitudes are within the reported experimental uncertainty. The MAE values for both are smaller than the measurement uncertainty associated with the overall isentropic efficiency $(< \pm 0.03)$ [7]. The proposed correlation shows a larger mean absolute error compared to the Doughty et al. correlation from Equation 1 but has fewer coefficients to fit. It is important to note that while the AHRI-540 polynomial correlation serves as the ground truth in this context, the proprietary measurement data used to derive these polynomials (e.g., from Bitzer) is subject to a permitted rating uncertainty of up to 5% (in standard regions) or 10% (in extended regions) according to AHRI-540 rating uncertainty limits.

Table 4: Quantitative error metrics for the isentropic efficiency correlations (compared to measured data from [8]):

	Mean Absolute Error [-]	Maximum Absolute Error [-]
Own Correlation	0.0212	0.1920
Doughty et al.	0.0120	0.0580

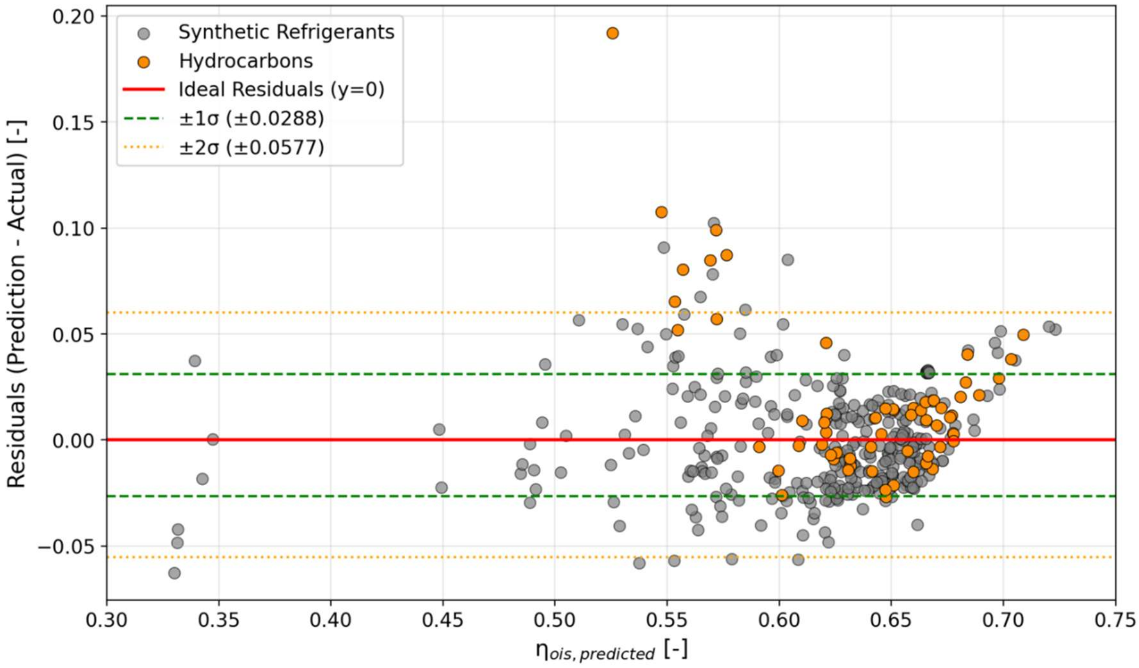

Figure 2: Residual plot of Equation 3 with the experimental dataset from [7]

While Table 4 summarizes the aggregated error metrics, a more detailed view of the correlation performance across the entire dataset is provided by the residual plot in Figure 2. The plot shows the residuals for each data point as a function of the predicted isentropic efficiency $\eta_{ois,predicted}$. Sigma (σ) represents the residual standard deviation, quantifying the typical scatter of the predictions relative to the experimental data. The $\pm 1\sigma$ and $\pm 2\sigma$ intervals serve as statistical thresholds to evaluate correlation precision and identify significant outliers. It can be observed that the majority of the data points are scattered randomly around the zero-error line and lie within the $\pm 2\sigma$ error band (orange dashed lines), with a high concentration within the $\pm 1\sigma$ band (green dashed lines). However, the plot also reveals some systematic trends. For instance, at higher predicted efficiencies (e.g., $\eta_{ois,predicted} > 0.65$), the correlation tends to slightly overpredict the efficiency (residuals are predominantly positive). Furthermore, a cluster of points shows larger positive residuals in the range of $\eta_{ois,predicted}$ between 0.50 and 0.65, indicating a tendency to overestimate the isentropic efficiency. The correlation also exhibits a relatively poor fit in regions of low isentropic efficiency. Specifically, in the efficiency range of approximately 0.30 to 0.35, the correlation struggles to capture the underlying physics, leading to discrepancies that can correspond to a relative error of 15% or more.

3.3 Analysis of Operating Range and Limitations

To further validate the accuracy of the correlation across the entire range of the compressor envelope, the 10-coefficient polynomial correlation according to AHRI-540 serves as the benchmark. The coefficients for this correlation and operating limits were obtained via the Bitzer Tool [11] with a constant superheat of 10 K. For comparison the isentropic efficiency was calculated with Equation 4 with calculated states at suction and discharge with CoolProp 7.0. [5].

$$\eta_{ois} = \frac{P_{is}}{P_{el}} = \frac{\dot{m}\cdot(h_{2is}-h_1)}{P_{el}} \tag{4}$$

A qualitative comparison of the correlation performance is presented in Figures 3, 4 and 5, which display the isentropic efficiency contour plots for R1234yf and R290 with a constant superheat of 10 K, respectively. Figure 3 shows the benchmark data for the reciprocating compressor for R1234yf (Model 2DES-3Y) and for R290 (Model 2DESP-3Z) according to the 10-coefficient polynomial derived from the Bitzer Tool. The predictions from the proposed correlation are shown (Equation 3) in Figure 4, and the predictions from the Doughty et al. correlation from Equation 1 in Figure 5.

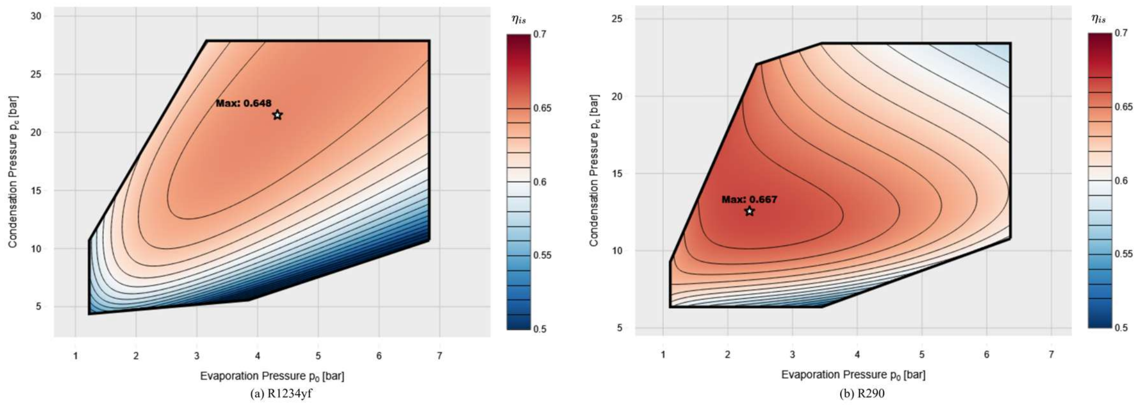

Figure 3: Contour plots for η_{ois} for (a) R1234yf and (b) R290 with AHRI 540 10-coefficient polynomial [11]

A direct comparison of Figure 3(a) and 3(b) reveals significant topological differences in the compressor's performance with each fluid. The map for R1234yf shows a well-defined area of maximum efficiency $\eta_{ois,max}$ located at a medium evaporation pressure and a high condensation pressure. On the contrary, the map for R290 exhibits a fundamentally different performance landscape, with a higher maximum efficiency $\eta_{ois,max}$ located in a completely different operating region: at low evaporation pressures and lower condensation pressures.

However, the unconventional topology of the isolines in Figure 3(b) suggests that these results must be interpreted with caution. The AHRI 540 10-coefficient polynomial correlation is highly sensitive to the distribution of underlying experimental data points [12]. The observed efficiency increases at higher condensation pressures while maintaining lower evaporation pressures for R290 likely stems from mathematical artifacts inherent to the high-order polynomial fit rather than the physical compressor behavior. Therefore, while the Figure 3(b) reflects the mathematical regression of the available dataset for the R290 compressor, the resulting gradients in certain regions may not fully represent the actual isentropic efficiency of the compressor, but rather emphasize the limitations of empirical modeling with the AHRI 540 10-coefficient polynomial.

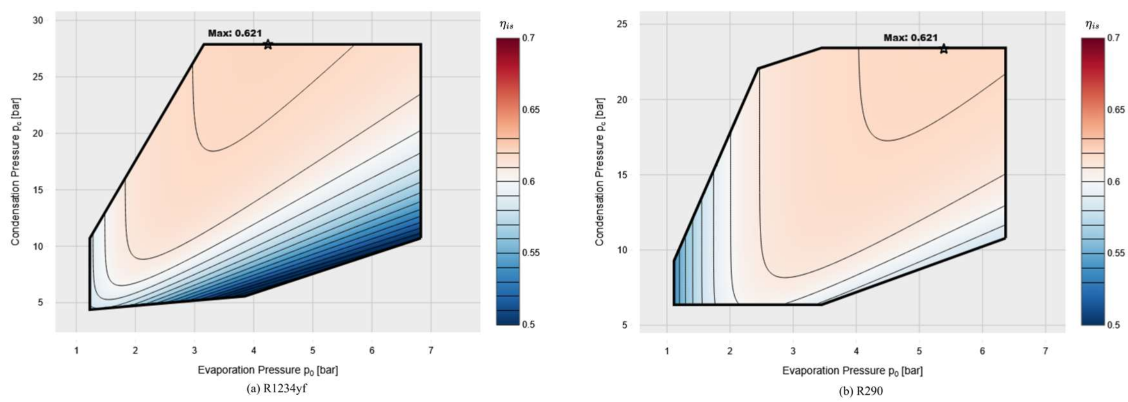

Figure 4: Contour plots for η_{ois} for (a) R1234yf and (b) R290 with correlation from Equation 3

However, these topological differences in the compressor's performance between the fluids cannot be captured by either from the proposed correlation or the correlation from O. Doughty et. al (by comparing Figure 4 and 5). They show a high degree of similarity between the plots for the two refrigerants. Both (a) and (b) predict an identical maximum efficiency $\eta_{ois,max}$ and display a nearly identical topological structure. In both cases, the correlation from O. Doughty et. al predicts a single area of high efficiency located in the region of moderate to high evaporation pressure and the highest possible condensation pressure. This result contrasts sharply with the benchmark AHRI-540 10 coefficient polynomial (shown previously in Figure 3), which demonstrated fundamentally different performance landscapes for the two fluids. While the proposed correlation (Figure 4a) successfully captures the general topology for R1234yf at medium to high evaporation pressures, it fails to reproduce the distinct behavior of R290. The observed deviations can be attributed to an insufficient experimental database for the R290 compressor used for the 10-coefficient polynomial regression. Therefore, the benchmark map itself may lack the necessary physical fidelity in certain operating regions, complicating a direct comparison with the proposed model. Independent of these benchmark-related inaccuracies, the current challenges in capturing the distinct topology behavior of the isentropic efficiency are hypothesized to stem from two primary factors. First, the SR algorithm, which minimizes with MSE, is statistically biased by the data imbalance visible in the dataset distribution (Figure 1), which is heavily dominated by synthetic refrigerants. Second, and more fundamentally, the inputs of the correlation are likely insufficient to describe the distinct physics; the experimental dataset explicitly used different oils (POE for synthetics, PAG for HCs), a critical, unmodeled variable that has probably a major effect on efficiency. This highlights a limitation and a direction for future refinement, to improve the generalizability of the proposed correlation across these distinct refrigerant classes.

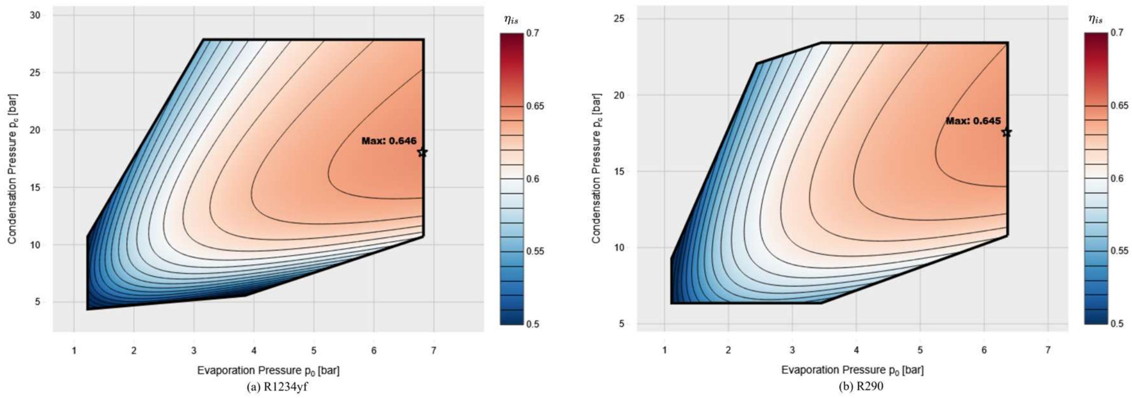

Figure 5: Contour plots for η_{ois} for (a) R1234yf and (b) R290 with correlation from Doughty et al. [3]

A qualitative analysis of the results from Doughty et al. (Figure 5) reveals a performance landscape that is distinct from both the AHRI benchmark and the correlation from Equation 3. In both cases, the correlation locates the peak efficiency region at high evaporation pressures and medium-to-high condensation pressures. While the R_x coefficient introduces minor changes to the contour gradients, particularly at low evaporation pressures where the efficiency for R290 is predicted to be lower than for R1234yf, the overall topology remains fundamentally the same for both fluids. When compared to the AHRI benchmark (Figure 3a), the Doughty et al. correlation (Figure 5a) captures the general characteristic for R1234yf, although it shifts the peak efficiency to a higher evaporation pressure region than the benchmark. In contrast to the predictions of Equations 1 and 3, the benchmark data demonstrates a peak efficiency region at low evaporation pressures for R290. A physics-based correction term could improve the topology. Another difference between the proposed correlation and the 10-coefficient polynomial is that the required experimental dataset for the proposed correlation is fundamentally different from that of the AHRI 10-coefficient polynomial. The polynomial requires a dense grid of operating points for a single fluid. In contrast, the proposed correlation, which uses properties like the acentric factor ω and compressibility factor at suction state Z_{suc} as inputs. Consequently, it requires a diverse, multi-fluid dataset to capture the variance in these thermophysical properties.

4. Conclusion and future work

This work successfully demonstrated a novel methodology for developing a refrigerant-generalized compressor performance correlation using symbolic regression (SR). By departing from traditional, fluid-specific correlations like the AHRI-540 10-coefficient polynomial and instead using fundamental thermophysical properties as inputs, specifically the acentric factor ω, compressibility factor at suction Z_{suc} and critical temperature T_{crit} this approach moves toward a more physically interpretable correlation.

The SR framework (PySR) successfully identified a single, parsimonious correlation (Equation 3) from an experimental dataset, providing a good prediction accuracy for compressor isentropic efficiency across multiple refrigerant classes (HFOs, HFCs, HCs, and mixtures). Although the quantitative evaluation reveals inferior performance of the proposed correlation compared to the semi-empirical correlation from Doughty et al. from the literature, it is characterized by greater parsimony, reflected in a significantly reduced number of coefficients. However, a key limitation identified in this study is the proposed correlations current inability to capture the distinct topological shifts observed in some benchmark datasets, particularly the unconventional efficiency landscape of R290. While Equation 3 successfully aligns with the general performance trends of synthetic refrigerants, it does not yet replicate the specific gradients of hydrocarbons at low evaporation pressures. It must be noted, however, that the AHRI-540 benchmark for R290 itself likely contains mathematical artifacts due to an insufficient experimental database in peripheral operating regions. Thus, the proposed correlations failure to replicate these artifacts may reflect a more stable mathematical structure.

Table 5 summarizes the findings by comparing the parametrization and complexity, performance and limitations, and the resulting utility and added value of each approach.

Table 5. Overview of correlations characteristics

Correlation	Parametrization & Complexity	Performance & Limitations	Utility & Added Value
AHRI-540 Polynomial	High complexity (10 fitted coefficients). Inputs restricted to operating temperatures only (t_0, t_c).	High local accuracy, but invalid for extrapolation. Requires dense experimental grids for every fluid.	Established industry standard for rating specific operating points but lacks transferability.
Doughty et al. Semi-Empirical	Moderate complexity (5 fitted coefficients). Inputs include pressure ratio (π), superheat (ΔT_{sh}) and an empirical coefficient (R_x).	Relies on an empirical coefficient (R_x) to distinguish between synthetic and hydrocarbon fluids.	Improved data fitting over broad ranges Uses semi-empirical adjustments.
Proposed Correlation Semi-Empirical	Low complexity (2 fitted coefficients). Inputs are refrigerant dependent thermodynamic state variables $(Z_{suc}, \omega,\ T_{crit})$, superheat ΔT_{sh} and pressure ratio π	Generalized across fluid classes (HFC, HFO, HC) without fluid-specific tuning but with higher MAE than the Doughty et. al. correlation.	Establishes a correlation based on refrigerant dependent variables rather than empirical coefficients.

Future work must therefore focus on refining this correlation. The analysis suggests that a physics-based correction term is necessary to improve the accuracy and generalizability across these distinct refrigerant classes. In summary, it is possible to derive a refrigerant-generalized correlation for the given dataset with this physics-informed symbolic regression approach which establishes a valid and promising path toward truly generalized, interpretable compressor correlations, but further refinement of the equation is required to achieve the necessary topological accuracy. Unlike the AHRI 10-coefficient polynomial, which requires a dense data grid for a single fluid, the new generalized correlation needs a diverse, multi-fluid dataset to capture the variance in its thermophysical inputs like acentric factor and compressibility factor.

Nomenclature

Symbol	Description	Unit
a	Coefficient for correlation	-
k	Isentropic exponent	-
p	Absolute pressure	bar
P	Power	W
T	Temperature	K
Z	Compressibility factor	-
μ	Dynamic Viscosity	Pa s
π	Pressure Ratio	-
ρ	Density	kg/m³
ω	Acentric factor	-
ΔT_{sh}	Superheat at suction	K
y	Mole fraction	-
Subscripts		
c	condensation	
crit	critical point	
dis	discharge	
el	electrical	
hc	hydrocarbons	
max	maximum	
m	mixture	
syn	synthetic	
0	evaporation	

References

[1] Ermel, C., Bianchi, M. V. A., and Schneider, P. S., 2023, “Energy Model of an Air Source Heat Pump to Explore Performance Improvements under Cold Conditions: A Python Framework: Preprint,” Renew. Energy.

[2] Marchante-Avellaneda, J., Navarro-Peris, E., and Shrestha, S. S., 2024, “Comparison and Selection of Experimental Designs for the Characterization of Scroll and Reciprocating Compressors: Adjustment of Polynomial Models with Compact Experimental Samples,” Appl. Therm. Eng., 241, p. 122341. https://doi.org/10.1016/j.applthermaleng.2024.122341.

[3] Cranmer, M., 2023, “Interpretable Machine Learning for Science with PySR and SymbolicRegression.Jl.” https://doi.org/10.48550/arXiv.2305.01582.

[4] Montanez-Barrera, J. A., Barroso-Maldonado, J. M., Bedoya-Santacruz, A. F., and Mota-Babiloni, A., 2022, “Correlated-Informed Neural Networks: A New Machine Learning Framework to Predict Pressure Drop in Micro-Channels,” Int. J. Heat Mass Transf., 194, p. 123017. https://doi.org/10.1016/j.ijheatmasstransfer.2022.123017.

[5] Bell, I. H., Wronski, J., Quoilin, S., and Lemort, V., 2014, “Pure and Pseudo-Pure Fluid Thermophysical Property Evaluation and the Open-Source Thermophysical Property Library CoolProp,” Ind. Eng. Chem. Res., **53**(6), pp. 2498–2508. https://doi.org/10.1021/ie4033999.

[6] Lemmon, E. W., Bell, I. H., Huber, M. L., and McLinden, M. O., 2018, “NIST Standard Reference Database 23: Reference Fluid Thermodynamic and Transport Properties-REFPROP, Version 10.0, National Institute of Standards and Technology.” https://doi.org/10.18434/T4/1502528.

[7] Brendel, L. P. M., Bernal, S. N., Roskosch, D., Arpagaus, C., Bardow, A., and Bertsch, S. S., 2024, “Compressor Performance for Varying Compositions of High-Glide Mixtures R1233zd(E)/R1234yf and R1336mzz(Z)/R1234yf,” *13th International Conference on Compressors and Their Systems*, M. Read, S. Rane, I. Ivkovic-Kihic, and A. Kovacevic, eds., Springer Nature Switzerland, Cham, pp. 753–764. https://doi.org/10.1007/978-3-031-42663-6_60.

[8] Doughty, O., Arpagaus, C., Bertsch, S. S., and Brendel, L. P. M., 2024, “Set of Performance Correlations for Reciprocating Compressor Covering Synthetic and Hydrocarbon Refrigerants.”

[9] Poling, B. E., Prausnitz, J. M., and O’Connell, J. P., 2020, *Properties of Gases and Liquids, Fifth Edition*, McGraw-Hill Education, New York, N.Y.

[10] Atkins, P., Paula, J. D., and Keeler, J., 2022, *Atkins’ Physical Chemistry*, Oxford University Press. https://doi.org/10.1093/hesc/9780198847816.001.0001.

[11] “Bitzer Websoftware.” [Online]. Available: www.bitzer.de/websoftware/.

[12] Marchante-Avellaneda, J., Corberan, J. M., Navarro-Peris, E., and Shrestha, S. S., 2023, “A Critical Analysis of the AHRI Polynomials for Scroll Compressor Characterization,” Appl. Therm. Eng., 219, p. 119432. https://doi.org/10.1016/j.applthermaleng.2022.119432.