Received: 1 June 2021 | Revised: 17 September 2021 | Accepted: 24 October 2021

DOI: 10.1002/aic.17518

THERMODYNAMICS AND MOLECULAR-SCALE PHENOMENA

AIChE JOURNAL

Use of 300,000 pseudo-experimental data over 1800 pure fluids to assess the performance of four cubic equations of state: SRK, PR, *tc*-RK, and *tc*-PR

Andrés Piña-Martinez | **Romain Privat** | **Jean-Noël Jaubert**

École Nationale Supérieure des Industries Chimiques, Laboratoire Réactions et Génie des Procédés (UMR CNRS 7274), Université de Lorraine, Nancy, France

Correspondence
Romain Privat and Jean-Noël Jaubert, École Nationale Supérieure des Industries Chimiques, Laboratoire Réactions et Génie des Procédés (UMR CNRS 7274), Université de Lorraine, 1 rue Grandville, 54000 Nancy, France.
Email: romain.privat@univ-lorraine.fr and jean-noel.jaubert@univ-lorraine.fr

Abstract

The guidelines to constitute a database containing, for 1800 pure fluids, more than 300,000 pseudo-experimental vapor pressure, liquid density, enthalpy of vaporization, and liquid heat capacity data are described. Such a database provides a tool for a fair assessment of the performances of equations of state to correlate (or predict) the properties of pure components. In this paper, we use it to assess the performance of four cubic equations of state (CEoSs): original Soave–Redlich–Kwong (SRK), original Peng–Robinson (PR), and the *translated-consistent* versions of PR (*tc*-PR) and RK (*tc*-RK). The deviations between calculated and experimental data are compared for the four CEoSs, which make it possible to discuss the influence of the α-function on the calculated vapor–liquid equilibrium properties and how the volume translation impacts the calculated volumetric properties. Eventually, the 1800 pure components are divided into 1252 non-self-associating and 548 self-associating compounds, and the accuracy of the CEoSs depending on the associating character of the molecules is discussed.

KEYWORDS
cubic equation of state, database, enthalpy of vaporization, liquid density, liquid heat capacity, Peng–Robinson, pure component, Soave–Redlich–Kwong, vapor pressure

1 | INTRODUCTION

Since the appearance of the first process flowsheet simulators, different physical-property databanks as well as many thermodynamic models have been incorporated in their brochures. Developers often claim that their models perform well in this or that region, can be reliably extrapolated, and are suitable for certain systems or applications. However, it is difficult to have a fair overview of the capabilities and limitations of a given thermodynamic model. Thus, it would be pertinent to define a common basis for comparison. At this stage, the following questions appear: "Which type of data are available for many fluids?" "Which type of properties are considered as essential?" "How could a thermodynamic model be evaluated quantitatively?"

These questions have been addressed by Jaubert et al.[1] in their work on the development of a high-quality reference database containing binary-system data for the cross-comparison of thermodynamic models and the assessment of their accuracies. With such a database, it is possible to determine how well a thermodynamic model predicts the phase-equilibrium properties of binary mixtures (vapor–liquid equilibrium [VLE], liquid–liquid equilibrium, vapor–liquid–liquid equilibrium, property change on mixing,[2] azeotropic and critical points) according to the associating character (AC) of their components. Moreover, it could be used to assess the impact of the selected mixing rules and the number

of parameters included within. Models such as the Peng–Robinson (PR) equation of state (EoS) armed with classical or advanced mixing rules, as well as the PC-SAFT EoS with classical mixing rules, no induced association scheme, and no regressed BIPs have been graded following the proposed procedure recently.[1,3,4]

Nevertheless, as stated by Agarwal et al.,[5,6] no matter how sophisticated the thermodynamic model may be or the number of parameters the mixture model may have, if the model is not even capable of representing accurately the behavior of the pure fluids contained in the mixture, there is no way to think about accurate and reliable mixture calculations. (As a convincing illustration, the reader is referred to fig. 3 in Agarwal et al.[6])

Keeping this in mind, it seems pertinent to assess the accuracy of a thermodynamic model for pure fluids prior to its extension to mixtures. Therefore, a standard basis of comparison is required. This is the reason why one of the primary motivations of this article is to serve as a reference for the objective assessment of the performances of thermodynamic models for pure fluids. To reach this objective, this article presents the guidelines that were used for constituting a reference database. The data collected for the proposed database are a fraction of those available in the database commercialized by the Design Institute for Physical Properties (DIPPR) and includes experimental vapor pressure, liquid density, enthalpy of vaporization, liquid heat capacity, as well as critical temperature, pressure, and volume data for 1800 pure fluids. It is believed that such a database, which contains more than 300,000 experimental data points, is the perfect tool to identify the strengths and weaknesses of a given thermodynamic model for pure compounds.

Additionally, the deviations between the calculated and experimental data are determined for the >300,000 data points available in the proposed database for the original Soave–Redlich–Kwong (SRK),[7] original PR,[8] *tc*-RK, and *tc*-PR EoSs[9] in order to assess their accuracies.

This article is organized into two sections. In the first one, a detailed description of how the database was built is presented. This includes the choice of key properties, molecules, and temperature ranges, resulting in more than 300,000 data points, which will be used to assess the accuracy of any thermodynamic model for pure components. In the following section, the four CEoSs benchmarked in this study are briefly introduced. Their capacity to reproduce experimental data is discussed then. Because it is often said that associating molecules cannot be properly modeled from EoSs that are not based on a theory accounting specifically for association bonding, the discussion is oriented toward the relation between the AC of the studied molecules and the capacity of EoSs to model their behavior.

2 | DESCRIPTION OF THE GUIDELINES TO CONSTITUTE A REFERENCE DATABASE OF PURE COMPONENTS

The objective of this section is to explain our scientific approach and answer the following questions:

1. How can the existing commercial databases help us to reach our objective (i.e., to define a reference database for evaluating the performance of thermodynamic models to correlate (or predict) the properties of pure components)?
2. What is the minimum list of properties that must be available for a pure species to be included in the reference database?
3. How many molecules should (and could) be included in the reference database of pure components?
4. How to define the temperature ranges in which deviations should be calculated?

2.1 | Existing databases of pure fluids

Concerning experimental data on pure fluids, three databases exist as the most used by researchers working in the field of chemical engineering thermodynamics. The first is the DIPPR database, which contains recommended values and accuracy estimates for 34 constant properties and 15 temperature-dependent properties spanning 2230 compounds. Another database is the Dortmund Data Bank (http://www.ddbst.com/ddb.html), which stores the largest amount of data worldwide for the thermophysical properties of pure components and their mixtures. The third one is the ThermoData Engine database of the National Institute of Standards and Technology, which provides experimental data for nearly 24,000 pure fluids and includes several methods for evaluation of the uncertainty and thermodynamic consistency of such data.

Since trustworthy data are contained in the aforementioned databases, any of them may be extremely useful for the definition of model comparison guidelines. Nevertheless, it is worth noting that an important feature of the DIPPR is that, apart from the experimental data points of the 15 temperature-dependent (*T*-dependent) properties, it proposes a fitting equation used for correlating *T*-dependent property data, reports the corresponding coefficients, and provides an evaluation of the mean error between predictions from the fitting equation and raw data. This enables the generation of pseudo-experimental data that are evenly distributed over wide ranges of temperature, which is desirable for the calculation of property deviations. Considering these features, the 12.3.0 version (May 2020) of the DIPPR database is used in this article as the starting point to build a reference database aimed at assessing the accuracy of EoSs for pure fluids.

2.2 | Key properties

The definition of key properties imposes a first criterion on the selection of the molecules that should be included in the assessment of a given thermodynamic model. The choice of such properties is based on an overview of the industrial needs but must also include the properties required to parameterize the EoS.

First of all, as stated by Agarwal et al.,[6] vapor pressure (P^{sat}) cannot be overlooked. Indeed, "regardless of the sophistication of your

thermodynamic model and the number of parameters in the mixing rule, you are in trouble if the vapor pressures are inaccurate". Such a property, which defines the endpoints of isothermal (*P*,*x*,*y*) phase diagrams, is thus of vital importance in the industry and, as was shown recently, is mandatory for reliable parameterizations of cubic and SAFT-type EoSs.[10,11]

The prediction of volumetric properties such as liquid density (ρ_{liq}^{sat}) is very important for the proper design of pipelines, pumps, or even heat exchangers. Additionally, Ramírez-Vélez et al.[11] found that parameterization of SAFT-type EoS can be performed reliably only if both liquid density and vapor pressure data are used. In the case of volume-translated CEoSs, it is necessary to have at least one experimental density data point at reduced temperature between 0.7 and 0.8 to properly determine the volume-translation parameter.

Properties such as enthalpies of vaporization ($\Delta_{vap}H$) and liquid heat capacities ($c_{P,liq}^{sat}$) are the key for the construction of energy balances. Inaccurate values of such properties may lead to unreliable designs of heat exchangers, expanders, or compressors, for instance. Finally, the reproduction of pure-component critical properties (especially T_c, P_c, and to a lesser extent v_c) is essential for an accurate representation of binary critical lines[12] (made up of binary critical points).

The ideal scenario would be to assess model performances considering molecules for which all the properties of interest (P^{sat}, ρ_{liq}^{sat}, $\Delta_{vap}H$, and $c_{P,liq}^{sat}$) are available. Nevertheless, due to a lack of experimental data, such a choice would drastically limit the number of pure components that could be embedded in the pure-component reference database we plan to build. In consequence, it is necessary to define the indispensable properties that should be available for any molecule included in the proposed database. Such properties are vapor pressure and liquid density; the reasons of this choice are as follows:

1. Satisfactory accuracy on the reproduction of P^{sat} and ρ_{liq}^{sat} is the minimum requirement for the industry and computer-aided product design developers.
2. In the case of CEoSs, to make an α-function *consistent*, the corresponding parameters must be fitted to P^{sat} data, at least.[10] Moreover, if the volume-translation concept is applied, at least one data point of the $\rho_{liq}^{sat}(T)$ curve is required.
3. In the case of SAFT-type EoS, Ramírez-Vélez et al.[11] concluded that it is necessary and sufficient to fit the PC-SAFT EoS parameters to both P^{sat} and ρ_{liq}^{sat} data.

In summary, molecules included in the database should have accurate data for P^{sat} and ρ_{liq}^{sat}; when available, $\Delta_{vap}H$ and $c_{P,liq}^{sat}$ must also be included.

It is worth noting that the EoS can predict only residual properties. Consequently, for estimating the liquid heat capacity, it is necessary to add an ideal-gas heat capacity term ($c_{P,IG}$) to the residual heat capacity. Therefore, an accurate correlation for the $c_{P,IG}(T)$ function is also needed when accurate $c_{P,liq}^{sat}$ data are available.

2.3 | Assessment of correlation accuracy

In the previous section, it has been stated which experimental properties should be available in order to incorporate a pure component in the proposed database. However, even if temperature-dependent correlations are available in the DIPPR database, it is necessary to guarantee their reliability. This is a key issue since, as Agarwal et al.[6] wrote: "one of the most heartbreaking experiences for a thermodynamicist is to watch someone spending a large effort refining a model with far more precision than the basic data allows."

2.3.1 | Presentation of the three types of errors

In order to evaluate whether for a given molecule this or that property correlation is accurate, a strategy inspired by the work of Piña-Martinez et al.[10] has been followed. For each correlation, three types of error were defined:

- *Type I error*: error reported in the DIPPR. Such an error is set by the DIPPR developers to the maximum uncertainty among the data series used to determine the coefficients of a correlation.
- *Type II error*: actual average uncertainty. Such an average error (not reported in the DIPPR database) is calculated from the knowledge of the uncertainty associated with each data series that was used to determine the coefficients of a correlation. For a given property and a given pure fluid, it is determined as

$$\text{Type II error} = \frac{\sum_{j=1}^{n_{series}} N_{data,j} \times \text{Error}_j}{\sum_{j=1}^{n_{series}} N_{data,j}} \quad (1)$$

where $N_{data,j}$ is the number of experimental data points available in the series j, and Error_j is the corresponding experimental error (as reported in the DIPPR database).

- *Type III error*: mean absolute percentage error (MAPE) over 50 points between pseudo-experimental data (generated by a DIPPR correlation) and calculated values returned by the generalized version of the *tc*-PR model (the acronym of which is *gen-tc*-PR). It is evaluated according to Equation (2):

$$\text{Type III error} = \frac{100}{N_T} \sum_{i=1}^{N_T} \left| \frac{X_{i,exp} - X_{i,calc}}{X_{i,exp}} \right| \quad (2)$$

$X_{i,exp}$ and $X_{i,calc}$ correspond to the pseudo-experimental and calculated values of the property X. In this article, $N_T = 50$ (the 50 points are evenly distributed between T_{min} and T_{max} that define the temperature range in which the correlation is valid).

Type III error was introduced because, from our experience with the DIPPR database, we believe that type-I and type-II errors are often overestimated. Consequently, many correlations would be disregarded if only those for which type-I or type-II errors are less than 5% would be kept. For this reason, but also to detect possibly typing errors in the coefficients of the correlations or in the reported uncertainties, our idea was to use a predictive EoS whose accuracy has been attested on many components to define a type-III error. In this article, an updated version of the predictive *gen-tc*-PR EoS (denoted as *gen-tc*-PR 2020 hereafter) is used. Such a model,[9,10] which requires only the knowledge of $T_{c,\text{exp}}$, $P_{c,\text{exp}}$, and ω_{exp} writes as

$$P(T,v)=\frac{RT}{(v+c)-b}-\frac{a_c\cdot\alpha(T_r)}{(v+c)(v+c+b)+b(v+c-b)} \text{ with}$$

$$:\begin{cases}\alpha(T_r)=T_r^{2(M-1)}\exp\left[L\left(1-T_r^{2M}\right)\right]\\ c=v_{\text{liq}}^{\text{sat},u-PR}(T_r=0.8)-v_{\text{liq,exp}}^{\text{sat}}(T_r=0.8)\\ \eta_c=\left[1+\sqrt[3]{4-2\sqrt{2}}+\sqrt[3]{4+2\sqrt{2}}\right]^{-1}\approx 0.25308\\ a_c=\dfrac{40\eta_c+8}{49-37\eta_c}\dfrac{R^2T_{c,\text{exp}}^2}{P_{c,\text{exp}}}\approx 0.45724\dfrac{R^2T_{c,\text{exp}}^2}{P_{c,\text{exp}}}\\ b=\dfrac{\eta_c}{\eta_c+3}\dfrac{RT_{c,\text{exp}}}{P_{c,\text{exp}}}\approx 0.07780\dfrac{RT_{c,\text{exp}}}{P_{c,\text{exp}}}\end{cases} \quad (3)$$

The ω-dependent correlations updated in this article are

$$\begin{cases}L=0.0544+0.7536\omega_{\text{exp}}+0.0297\omega_{\text{exp}}^2\\ M=0.8678-0.1785\omega_{\text{exp}}+0.1401\omega_{\text{exp}}^2\end{cases} \quad (4)$$

In Table 1, the MAPE and the standard deviation (SD) obtained with the *gen-tc*-PR 2020 EoS over 1786 molecules are reported. Such values are important for the acceptance test presented hereafter.

2.3.2 | Correlation acceptance tests

A correlation is declared as acceptable if it passes *both tests* described in this section.

Test 1: The type III error of a given property correlation $X\left(X\in\left\{P^{\text{sat}},\rho_{\text{liq}}^{\text{sat}},\Delta_{\text{vap}}H,c_{P,\text{liq}}^{\text{sat}}\right\}\right)$ is less than the threshold of acceptance (noted TH_X) defined as follows:

$$TH_X^{\text{test1}}=\text{MAPE}_X+6\text{SD}_X \quad (5)$$

where MAPE_X (for mean average percent error) and SD_X (standard deviation) of property X are obtained from Table 1.

This test aims at detecting possibly typing errors in the coefficients of the correlations or in the values of uncertainties reported in the DIPPR database. It also aims at disregarding correlations that lead to totally unexpected values so that the properties of such a component cannot be correlated by a classical EoS (this usually happens for a few inorganic molecules). In Equation (5), it was decided to multiply SD_X by 6 because it is believed that such a value makes it possible to detect typing errors among the data and to exclude undesired molecules (such as metals), but not to reject data that are poorly correlated by the *gen-tc*-PR 2020 EoS. To be concrete, let us take an example. Table 1 reports that the average deviation on P^{sat}, calculated by the *gen-tc*-PR 2020 EoS, is 1.7% (and $\text{SD}_{P^{\text{sat}}}=1.5\%$). If, for a given molecule, the average deviation between pseudo-experimental data stemming from a DIPPR correlation and values calculated with the *gen-tc*-PR 2020 EoS is higher than 10.7% ($\text{MAPE}_{P^{\text{sat}}}+6\text{SD}_{P^{\text{sat}}}$), such a correlation is rejected.

Test 2: At least one out of the three types of error is less than or equal to its respective acceptance threshold. These values are reported in Table 2.

Once again, to be concrete, let us take an example. If for a given molecule

1. The error reported by the DIPPR for the correlation on P^{sat} is 6%;
2. The average DIPPR average for the correlation on P^{sat} is 5.1%; and
3. The *gen-tc*-PR 2020 EoS correlates the pseudo-P^{sat} data with an accuracy of 2% (less than $\text{MAPE}_{P^{\text{sat}}}+\text{SD}_{P^{\text{sat}}}$).

Then, the correlation is declared acceptable.

2.4 | Molecules to be incorporated in the database

The DIPPR provides constant properties and a set of T-dependent correlations for 2230 molecules. Among these (see filter 2 below), only the ones for which both key properties (P^{sat}, $\rho_{\text{liq}}^{\text{sat}}$) are available were kept. In addition, all molecules requiring a specific treatment (see filter 1 below) were disregarded. Therefore, a molecule was declared as suitable if it remained after application of both filters described below.

Filter 1: A molecule is removed if

- T_c is greater than 2000 K. It is considered that such a molecule cannot be accurately represented by an EoS routinely used in process simulators and relying on a corresponding-states principle;

TABLE 1 MAPE and SD obtained with the 2020 *generalized* version of the *tc*-PR model used for this article

	Property X			
	P^{sat}	$v_{\text{liq}}^{\text{sat}}$	$\Delta_{\text{vap}}H$	$c_{P,\text{liq}}^{\text{sat}}$
Number of molecules	1786	1786	1503	895
MAPE_X	1.7%	2.1%	2.6%	4.8%
SD_X	1.5%	1.3%	1.6%	4.3%

TABLE 2 Acceptance thresholds for the different property correlations and types of error

	Acceptance thresholds for each property correlation				
Error type	P^{sat}	v_{liq}^{sat}	$\Delta_{vap}H$	$c_{P,liq}^{sat}$	$c_{P,IG}$
Type I (error reported in the DIPPR database)	5%	5%	5%	5%	5%
Type II (average exp. error calculated by Equation 1)	5%	5%	5%	5%	5%
Type III (deviation calculated with the *gen-tc*-PR 2020 EoS by Equation 2)	$TH_X^{test\,2} = MAPE_X + SD_X$[a]				—

[a]$MAPE_X$ and SD_X are obtained from Table 1.

- it is banned by the DIPPR due to ambiguous data, self-accelerating decomposition, or instability;
- it is a quantum fluid (e.g., hydrogen, helium, etc.).

After this filter, 47 molecules were removed and 2183 remained.

Filter 2: A molecule is removed if one correlation for $P^{sat}(T)$ or $\rho_{liq}^{sat}(T)$ is missing or does not pass the acceptance tests presented in Section 2.3.2. After this filter, 383 molecules were removed and 1800 remained.

Note that after this last filter, 385 molecules should be removed actually. Among these molecules were *n*-dotriacontane and *n*-hexatriacontane, which exhibit an inaccurate correlation for $P^{sat}(T)$. Since *n*-dotriacontane and *n*-hexatriacontane are included in the benchmark database for binary systems published by Jaubert et al.,[1] it was decided not to remove them from the database.

2.4.1 | Associating character

The thermodynamic behavior of molecules showing AC, that is, involved in a hydrogen bond, is considered by most of developers (and SAFT EoS authors in particular) as more complicated than that of non-associating molecules. Thus, capturing the behavior of associating molecules appears as a challenge. As a consequence, how a model performs with non-self-associating (NSA) and self-associating (SA) molecules gives a deeper insight into its performance. In the case of binary systems, a major feature of the database proposed by Jaubert et al.[1] is that it provides a binary system classification according to the AC of their components. It is thus possible to have a measure of the complexity of modeling the thermodynamic properties of mixtures depending on the nature and strength of the association phenomena involved in the system.

In this article, the AC of each pure species is quantified following a methodology explained in Appendix A1. In short, the screening charge density distribution (sigma profile) of each molecule was generated and analyzed to determine the tendency of such a compound to associate. For the 1800 components involved in the proposed database, it results in 1252 NSA and 548 SA molecules.

2.5 | Temperature ranges and deviations

For a given pure component j and a given property X ($X \in \left\{P^{sat}, \rho_{liq}^{sat}, \Delta_{vap}H, c_{P,liq}^{sat}\right\}$), 50 equidistant pseudo-experimental data points are generated in their valid temperature range [T_{min}, T_{max}]. The temperature ranges for each property are mentioned in Supporting Information (Appendix A2). Concerning deviations, the MAPEs are evaluated according to Equation (6) as the average of the absolute percentage deviations between the experimental and calculated properties over N_T temperature points.

$$MAPE_X = \frac{100}{N_T}\sum_{i=1}^{N_T}\left|\frac{X_{j,exp}(T_i) - X_{j,calc}(T_i)}{X_{j,exp}(T_i)}\right| \quad (6)$$

$X_{j,exp}(T_i)$ and $X_{j,calc}(T_i)$ correspond to the calculated and experimental values of component j's property X at temperature T_i. In the present article, $N_T = 50$. It is worth noting that the first temperature (T_1) and the last temperature (T_{50}) of the range correspond exactly to T_{min} and T_{max}, respectively.

2.6 | Guidelines summary and data points

The guidelines presented in this article lead to a database containing 1800 molecules that have accurate correlations for P^{sat} and ρ_{liq}^{sat}. Among them, 1536 fluids possess an accurate correlation for $\Delta_{vap}H$ and 890 for $c_{P,liq}^{sat}$, resulting in a database containing 301,300 high-quality pseudo-experimental points. Moreover, critical properties (T_c, P_c, and v_c) are available for the 1800 molecules, which results in 5400 additional data points. In total, the proposed database contains 306,700 pure-component data points for for the assessment of thermodynamic models' performances. The name, the AC, and the critical properties (T_c, P_c, and v_c) of each molecule along with the temperature range and the identification (ID) number assigned by the DIPPR to each property correlation are reported in Appendix A3. In addition, Table 3 gives a short overview of such a database.

3 | BENCHMARK OF THE SRK, PR, *tc*-RK, AND *tc*-PR EoSs

In this section, the performances of four CEoS are evaluated by assessing their accuracy over the 306,700 pseudo-experimental data points available in the reference database for pure compounds developed in this article.

TABLE 3 Overview of the data available in the proposed database

Total number of pure components: 1800 (1252 non-self-associating + 548 self-associating). Total number of pseudo experimental data points: 3,06,700						
Number of components for which the property is available						
P^{sat}	ρ_{liq}^{sat}	$\Delta_{vap}H$	$c_{P,liq}^{sat}$ (and $c_{P,IG}$)	T_c	P_c	v_c
1800 (90,000 data points)	1800 (90,000 data points)	1536[a] (76,800 data points)	890[b] (44,500 data points)	1800 (1800 data points)	1800 (1800 data points)	1800 (1800 data points)
306,700 data points						

[a]Among the 1536 components, 1114 are non-self-associating and 422 are self-associating.
[b]Among the 890 components, 695 are non-self-associating and 195 are self-associating.

3.1 | Models

The first two CEoSs are the well-known SRK[7] and PR[8] EoSs. The other two models are the *translated*[13,14] and *consistent*[15,16] versions of the RK (*tc*-RK) and PR (*tc*-PR) EoSs.[9] The detailed expressions of the SRK, PR, *tc*-RK, and *tc*-PR EoSs are summarized in Table 4. It is worth recalling that for the *tc*-RK and *tc*-PR models, *translated* stands for the fact that a substance-dependent volume-translation parameter *c* has to be determined, component by component, in order that the translated EoS exactly reproduces the experimental saturated liquid volume at a reduced temperature of 0.8 [$v_{liq,exp}^{sat}(T_r = 0.8)$], that is

$$c = v_{liq}^{sat,u-CEoS}(T_r = 0.8) - v_{liq,exp}^{sat}(T_r = 0.8) \tag{7}$$

where $v_{liq}^{sat,u-CEoS}(T_r = 0.8)$ is the molar volume calculated with the original (*untranslated*) CEoS at $T_r = 0.8$. In turn, *consistent* stands for the fact that such EoSs are coupled with an α-function that passes the *consistency test* developed by Le Guennec et al.[9,15,16] and therefore guarantees safe property predictions in both the subcritical and supercritical domains. Such a consistency test is expressed as a list of constraints recalled in Equation (8); any α-function obeying these constraints is considered as consistent. This set of conditions has been called *consistency test for an α-function* in analogy with the consistency tests developed to certify the quality of experimental binary VLE data.

$$\text{For any reduced temperature } (T_r): \begin{cases} \alpha(T_r) \geq 0 \text{ and } \alpha \text{ continuous} \\ \dfrac{d\alpha}{dT_r}(T_r) \leq 0 \text{ and } \dfrac{d\alpha}{dT_r} \text{ continuous} \\ \dfrac{d^2\alpha}{dT_r^2}(T_r) \geq 0 \text{ and } \dfrac{d^2\alpha}{dT_r^2} \text{ continuous} \\ \dfrac{d^3\alpha}{dT_r^3}(T_r) \leq 0 \end{cases} \tag{8}$$

As shown in Table 4, the *tc*-RK and *tc*-PR EoSs are armed with the highly flexible three-parameter α-function proposed by Twu et al. in 1991[17]:

$$\alpha(T_r) = T_r^{N(M-1)} \exp\left[L\left(1 - T_r^{MN}\right)\right] \tag{9}$$

For the 1800 pure components available in the proposed database, the three *L*, *M*, and *N* parameters were determined for both the *tc*-RK and *tc*-PR EoSs following the fitting procedure that Piña-Martinez et al.[10] advised to implement. In short, the *L*, *M*, and *N* parameters were determined in order to reproduce as well as possible P^{sat} (at least), and/or $\Delta_{vap}H$, and/or $c_{P,liq}^{sat}$ (if available) experimental data and to guarantee that the α-function is *consistent*. The resulting parameters are reported in Appendix A4.

3.2 | Results and discussion

The MAPEs on $\left\{P^{sat}, v_{liq}^{sat}, \Delta_{vap}H, c_{P,liq}^{sat}, v_c\right\}$, as defined in Equation (6), are reported in Table 5 for the PR, SRK, *tc*-RK, and *tc*-PR EoSs. The details of deviations for each molecule are reported in Appendix A5.

From the α-function standpoint, it is possible to observe the huge and well-known impact of the choice of a generalized or component-dependent function. In the case of the PR EoS, when it is coupled with the Twu91 α-function (case of the *tc*-PR EoS) rather than with a Soave-type one (original PR EoS), MAPE on P^{sat} is divided by almost 3, MAPE on $\Delta_{vap}H$ is 1% lower, and MAPE on $c_{P,liq}^{sat}$ is almost divided by 3. Although the reductions are smaller, the same trends are observed for the RK EoS when coupled with a Soave-type or a Twu91 α-function. It is noticeable that the original SRK yields much better results on $c_{P,liq}^{sat}$ than the original PR.

Concerning the impact of the different volumetric functions included in the attractive term of the investigated CEoSs, the SRK EoS leads to larger deviations (19%) on v_{liq}^{sat} than the PR EoS (8%). In turn, the impact of a volume translation is enormous, since it is possible to observe a reduction from 19% (SRK) to 3.5% (*tc*-RK), and from 8.8% (PR) to 2.1% (*tc*-PR). It is worth recalling that, as a general trend, the SRK EoS predicts volumes that are much larger than the experimental ones. In turn, the volumes predicted by the PR EoS may be larger or smaller than the experimental ones depending on the studied fluid. In the latter case, negative values of the volume-translation parameter (*c*) are reported.

Table 5 allows us conclude that the *tc*-PR EoS with an average overall deviation of less than 2% over the 306,700 available

TABLE 4 Expressions of the SRK, PR, *tc*-RK, and *tc*-PR EoS

SRK	PR	*tc*-RK	*tc*-PR
$P(T,v)=\frac{RT}{v-b}-\frac{a_c\cdot\alpha(T_r)}{v(v+b)}$	$P(T,v)=\frac{RT}{v-b}-\frac{a_c\cdot\alpha(T_r)}{v(v+b)+b(v-b)}$	$P(T,v)=\frac{RT}{(v+c)-b}-\frac{a_c\cdot\alpha(T_r)}{(v+c)(v+c+b)}$	$P(T,v)=\frac{RT}{(v+c)-b}-\frac{a_c\cdot\alpha(T_r)}{(v+c)(v+c+b)+b(v+c-b)}$
α-Function			
$\alpha(T)=\left[1+m\left(1-\sqrt{T_r}\right)\right]^2$ with $T_r=T/T_c$		$\alpha(T_r)=T_r^{N(M-1)}\exp\left[L\left(1-T_r^{MN}\right)\right]$ (L, M, and N are component-dependent and specific to each EoS)	
$m=0.480+1.574\omega_{\text{exp}}-0.176\omega_{\text{exp}}^2$	$m=0.37464+1.54226\omega_{\text{exp}}-0.26992\omega_{\text{exp}}^2$		
Volume-translation parameter			
Nonapplicable		$c=v_{\text{liq}}^{\text{sat},u-EoS}(T_r=0.8)-v_{\text{liq,exp}}^{\text{sat}}(T_r=0.8)$ (c is component-dependent and specific to each EoS)	
Attractive parameter at the critical temperature			
$a_c=\Omega_a\frac{R^2T_{c,\text{exp}}^2}{P_{c,\text{exp}}}$			
Parameter b			
$b=\Omega_b\frac{RT_{c,\text{exp}}}{P_{c,\text{exp}}}$			
$\Omega_a=\frac{1}{9\left(\sqrt[3]{2}-1\right)}\approx0.42748$ $\Omega_b=\frac{\sqrt[3]{2}-1}{3}\approx0.08664$	$\Omega_a=\frac{8(5X+1)}{49-37X}\approx0.45724$ $\Omega_b=\frac{X}{X+3}\approx0.07780$ $X=\left[1+\sqrt[3]{4-2\sqrt{2}}+\sqrt[3]{4+2\sqrt{2}}\right]^{-1}\approx0.25308$	$\Omega_a=\frac{1}{9\left(\sqrt[3]{2}-1\right)}\approx0.42748$ $\Omega_b=\frac{\sqrt[3]{2}-1}{3}\approx0.08664$	$\Omega_a=\frac{8(5X+1)}{49-37X}\approx0.45724$ $\Omega_b=\frac{X}{X+3}\approx0.07780$ $X=\left[1+\sqrt[3]{4-2\sqrt{2}}+\sqrt[3]{4+2\sqrt{2}}\right]^{-1}\approx0.25308$

experimental data points is extremely accurate. It is probably not only the safest (the *consistent* character of the α-function of such an EoS ensures safe property predictions in both subcritical and supercritical domains) but also the most accurate cubic EoS available.

3.2.1 | Associating character

As previously mentioned, the 1800 pure fluids are divided into two groups: those that are SA and those that are not (NSA). MAPE on $\left\{P^{sat}, v_{liq}^{sat}, \Delta_{vap}H, c_{P,liq}^{sat}, v_c\right\}$ for the 1252 NSA and the 548 SA compounds are reported in Table 6. A first point is that, for the original SRK and PR EoSs, the difference in performance between NSA and SA compounds is noticeable for all the properties of interest. For the *translated* and *consistent* models, similar deviations are observed for SA and NSA compounds except for the $c_{P,liq}^{sat}$ property for which deviations double when switching from NSA to SA compounds. We however, believe that this difference is not very significant. A detailed analysis of the results obtained with the *tc*-PR EoS shows that the vast majority of SA compounds (around 70%) have deviations on $c_{P,liq}^{sat}$ that are less than 2%. Half of the SA components have even deviations on $c_{P,liq}^{sat}$ that are less than 1.5%. On the other hand, a small number of SA components have deviations higher than 20%, which inevitably increases the average deviation value for this category of components. Table 6 shows that for SA compounds, the average deviation on $c_{P,liq}^{sat}$ obtained with the *tc*-PR EoS is 4.2%, whereas the calculated SD is found to be as high as 3.8%. On the other hand, for NSA components the SD is close to zero. A simple comparison of the average deviations (between SA and NSA components) is thus not very relevant. Another weakness of the comparison of deviations between SA and NSA components is the small number of SA components for which experimental heat capacities are available (less than 200).

To have a deeper insight into the role played by association and to better understand the relationship between the ability of an EoS to accurately reproduce the properties of pure species and their AC, it is now decided to define a criterion for *quality modeling*. Using the latter, molecules will be labeled as *well* or *badly* represented, and it will be possible to analyze how the SA and NSA fluids distribute between the two categories. The proposed criterion of *modeling quality* compares the deviations obtained on P^{sat} and v_{liq}^{sat} for a given fluid with the mean deviations obtained for all the NSA fluids. For a given molecule, if deviations on P^{sat} and v_{liq}^{sat} slightly depart from the corresponding mean values for NSA fluids, the molecule is considered as well modeled. Otherwise, it falls in the "badly modeled" category. Deviations on P^{sat} and v_{liq}^{sat} were retained because such properties are available for the 1800 molecules and, as previously discussed, are considered as key properties. The mean reference deviations on P^{sat} and v_{liq}^{sat} were calculated over the 1252 NSA components only, simply because the four studied CEoSs do not contain associating terms and

TABLE 5 MAPE on P^{sat}, v_{liq}^{sat}, $\Delta_{vap}H$, $c_{P,liq}^{sat}$, v_c, T_c, and P_c as predicted by the SRK, PR, *tc*-RK, and *tc*-PR EoS

EoS	MAPE on P^{sat} (1800 fluids)	MAPE on v_{liq}^{sat} (1800 fluids) ($T_r < 0.9$)	MAPE on $\Delta_{vap}H$ (1536 fluids)	MAPE on $c_{P,liq}^{sat}$ (890 fluids)	MAPE on v_c (1800 fluids)	MAPE on T_c and P_c[a] (1800 fluids)	Global average deviation over the 306,700 available data points
SRK	2.3%	19%	3.0%	5.5%	31%	0%	8.0%
PR	2.8%	8.8%	3.1%	7.3%	21%	0%	5.4%
tc-RK	1.2%	3.5%	2.0%	2.7%	24%	0%	2.4%
tc-PR	1.0%	2.1%	1.9%	2.5%	20%	0%	1.9%

[a]Such MAPEs are exactly zero since the four EoSs are parameterized in order to exactly reproduce $T_{c,exp}$ and $P_{c,exp}$.

TABLE 6 MAPE on P^{sat}, v_{liq}^{sat}, $\Delta_{vap}H$, $c_{P,liq}^{sat}$, and v_c depending on the associating character as predicted by the SRK, PR, *tc*-RK, and *tc*-PR EoS

EoS		MAPE on P^{sat} (1800 fluids)	MAPE on v_{liq}^{sat} (1800 fluids) ($T_r < 0.9$)	MAPE on $\Delta_{vap}H$ (1536 fluids)	MAPE on $c_{P,liq}^{sat}$ (890 fluids)	MAPE on v_c (1800 fluids)
SRK	NSA	2.1% (1252 fluids)	18% (1252 fluids)	2.8% (1114 fluids)	4.9% (695 fluids)	30% (1252 fluids)
	SA	2.9% (548 fluids)	21% (548 fluids)	3.5% (422 fluids)	7.6% (195 fluids)	33% (548 fluids)
PR	NSA	2.3%	7.7%	3.0%	6.8%	20%
	SA	4.0%	11%	3.5%	9.0%	23%
tc-RK	NSA	1.1%	3.6%	2.0%	2.3%	23%
	SA	1.4%	3.2%	1.9%	4.4%	25%
tc-PR	NSA	0.9%	2.0%	2.0%	2.1%	19%
	SA	1.2%	2.3%	1.8%	4.2%	21%

were initially intended to describe NSA components only. By only considering NSA compounds instead of all pure components of the database, a more severe test is proposed to include a molecule to the "well-modeled" category.

Quantitatively speaking, a pure component j will be considered as well modeled by a given EoS provided

$$\begin{cases} \left(\mathrm{MAPE}_{P^{sat}}^{EoS}\right)_j \leq \overline{\mathrm{MAPE}}_{P^{sat}}^{EoS,NSA} + \mathrm{SD}_{P^{sat}}^{EoS,NSA} \\ \text{and} \\ \left(\mathrm{MAPE}_{v_{liq}^{sat}}^{EoS}\right)_j \leq \overline{\mathrm{MAPE}}_{v_{liq}^{sat}}^{EoS,NSA} + \mathrm{SD}_{v_{liq}^{sat}}^{EoS,NSA} \end{cases} \quad (10)$$

where $\mathrm{SD}_X^{EoS,NSA}$ and $\overline{\mathrm{MAPE}}_X^{EoS,NSA}$ denote the mean SD and MAPE for property X calculated with a given EoS over the 1252 NSA components of the database. If one or both conditions of Equation (10) are not satisfied, the pure component is considered as "badly modeled."

Results are presented in Figures 1–4 and deserve discussion. A key point is that, regardless of the considered CEoS, about 75% of pure components are well represented, according to the criterion defined in Equation (10) (see the left panel in each figure). In addition, the database contains 70% (1252/1800) of NSA fluids and 30% of SA components. At first sight, it could be believed that the 75% well-modeled molecules match the NSA fluids plus some SA components. This is, however, not the case at all.

Another striking fact is that almost 67% of the SA molecules are well modeled (see the right panel in each figure), while almost 20% of the NSA compounds are badly modeled (refer to the middle panel in each figure). Once again, we might have been tempted to believe that all the NSA fluids would be well modeled by CEoS, but this is not the case. On the other hand, one would not have predicted that classical CEoSs reproduce satisfactorily more than half of the SA molecules without adding an association term.

In order to further analyze the results, we decided to investigate whether there was a correlation between the degree of hydrogen bonding and the accuracy with which the experimental data were correlated. In other words, we wanted to investigate whether the performance for short-chain associating molecules (where the weight of association is definitely very important) is less reliable or it follows similar trends as for SA compounds that possess a long organic chain, these long alkyl chains being likely to hide the effect of association. It was thus decided to split the 548 SA components into two categories: the short-chain and the long-chain components. We found that 125 molecules (among the 548) had a molecular weight less than or equal to that of *n*-pentanol and they were (somewhat arbitrary) classified as short-chain molecules. The remaining 423 molecules were classified as long-chain molecules. Calculations were performed with the *tc*-PR EoS, which made it possible to conclude that there is definitely no link between the strength of the hydrogen bond and the accuracy of the calculations. Indeed, 62% of the short-chain molecules and 68% of the long-chain molecules were found to be well modeled. The difference between these two percentages is small and not statistically significant. Water, methanol, and ethanol were declared badly modeled but many other short alcohols (propargyl alcohol, allyl alcohol, 2-methyl-2-propanol, 2-butanol, 1,2-propylene glycol, etc.) were declared well modeled. In addition, we observed some results for which it is difficult to find an explanation. As an example, considering the two isomers 1-butanol and 2-butanol, it is found that the first one is well modeled but the second one is not. Their structures are, however, highly similar.

Finally, a closer inspection across various chemical groups (defined by the DIPPR) is proposed.

Table 7 shows the distribution of well-modeled compounds for different chemical groups and for each EoS. It is possible to observe that the four EoSs have a ratio of well-modeled molecules greater than 65% and often close to 80% for alkanes, alkenes, alkynes, aromatics, esters, ethers, ketones, aldehydes, epoxides, peroxides, halogen compounds, amines, amides, sulfur and silicon compounds. Alcohols have 61% of well-modeled molecules except with the *tc*-PR model, which yields a ratio of 53%. Nitriles are modeled the worst, and this is related mostly to high deviations on v_{liq}^{sat}. Such large deviations on v_{liq}^{sat} can, however, not be ascribed to the strong AC of these compounds. A detailed analysis of the chemical structure of the 31 nitrile compounds available in the DIPPR database reveals that most of these molecules are

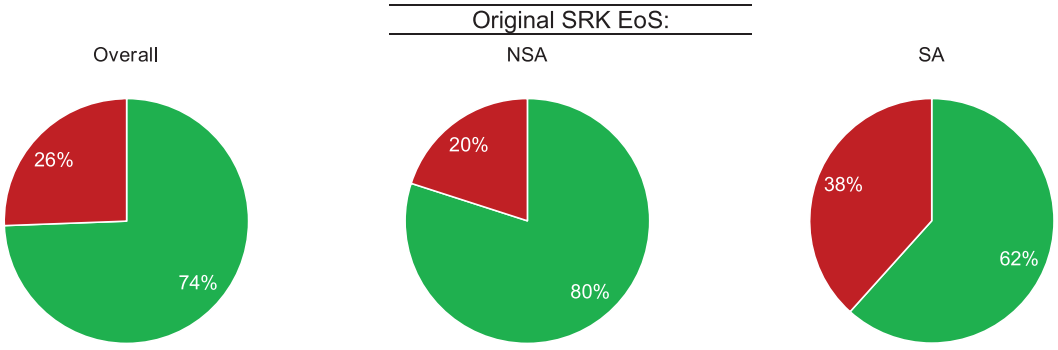

FIGURE 1 Ratio of well-modeled and badly modeled molecules by the SRK EoS classified into non-self-associating (NSA) and self-associating (SA) fluids. Green: well-modeled molecules. Red: badly modeled molecules

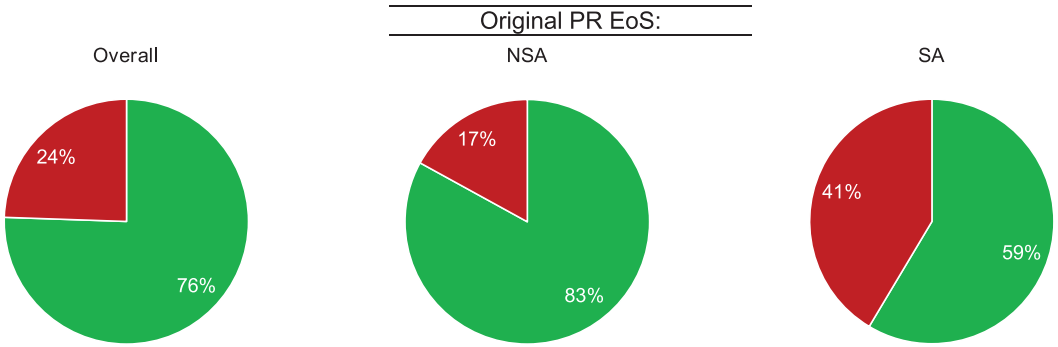

FIGURE 2 Ratio of well-modeled and badly modeled molecules by the PR EoS classified into non-self-associating (NSA) and self-associating (SA) fluids. Green: well-modeled molecules. Red: badly modeled molecules

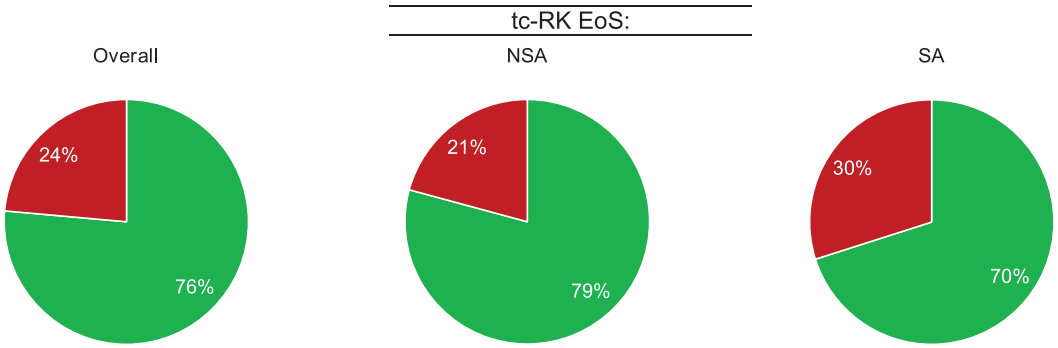

FIGURE 3 Ratio of well-modeled and badly modeled molecules by the *tc*-RK EoS classified into non-self-associating (NSA) and self-associating (SA) fluids. Green: well-modeled molecules. Red: badly modeled molecules

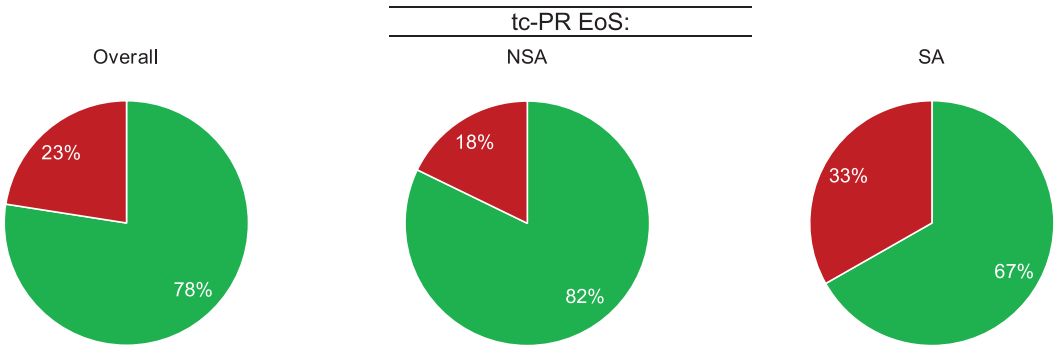

FIGURE 4 Ratio of well-modeled and badly modeled molecules by the *tc*-PR EoS classified into non-self-associating (NSA) and self-associating (SA) fluids. Green: well-modeled molecules. Red: badly modeled molecules

polyfunctional (in addition to the nitrile group, some contain double bonds, other contain a hydroxyl group, a NH_2 group, a methoxy group, etc.). It is believed that such a complexity explains the loss of accuracy that is observed. In turn, organic acids are not well modeled by the original SRK and PR EoSs while the *tc*-RK and *tc*-PR models achieve a ratio of 80% of well-modeled molecules.

To conclude this section, it can be highlighted that modern CEoSs like the *tc*-PR and *tc*-RK EoS have the capacity to model accurately around 70% of SA compounds and 80% of NSA compounds. We believe that this result counterbalances most of comments about the inadequacy of EoSs that do not include a specific association term in their formulation to model SA compounds.

TABLE 7 Distribution of well-modeled compounds across different chemical groups for the SRK, PR, *tc*-RK, and *tc*-PR EoS

Group	SRK (%)	PR (%)	*tc*-RK (%)	*tc*-PR (%)
Alkanes	88	90	85	94
Alkenes	92	90	84	87
Aromatics	80	85	87	83
Alkynes	83	83	78	83
Polyfunctional compounds	71	63	82	74
Esters/ethers	69	73	84	79
Organic acids	49	35	80	79
Inorganic compounds	63	67	63	60
Ketones/aldehydes	88	93	80	93
Nitrogen compounds	64	62	71	75
Alcohols	61	61	61	53
Other compounds	33	56	67	44
Organic salts	69	62	54	46
Epoxides	94	100	61	83
Peroxides	100	73	100	73
Halogen compounds	81	86	72	80
Nitriles	13	13	23	35
Amines/amides	70	70	72	68
Sulfur compounds	88	91	79	90
Silicon compounds	75	77	80	72

4 | CONCLUSION

In this article, we presented guidelines for the constitution of a database of pure-compound properties stemming from the DIPPR for the assessment of the performance of thermodynamic model for such compounds. These guidelines provide the number of molecules, the property data to be incorporated, and the temperature ranges between which pseudo-experimental points should be generated.

Application of such guidelines led to a database containing 1800 molecules for which, in addition to the critical coordinates (T_c, P_c, and v_c), accurate correlations for P^{sat} and ρ_{liq}^{sat} are available. Among them, 1536 fluids also possess an accurate correlation for $\Delta_{vap}H$ and 890 for $c_{P,liq}^{sat}$, resulting in a database containing 306,700 high-quality pseudo-experimental data points.

In addition, the accuracy of the SRK, PR, *tc*-RK, and *tc*-PR EoSs has been investigated by calculating the deviations between experimental data points embedded in the developed reference database and calculated values returned by the CEoSs. Results obtained over 1800 compounds and more than 306,000 data points highlight the following:

- The use of the highly flexible, component-dependent Twu91 α-function drastically improves the reproduction of vapor pressures, enthalpies of vaporization, and liquid heat capacities.
- There is a spectacular improvement in the reproduction of liquid molar volumes when a substance-dependent volume-translation parameter is used.
- The accuracy of the *tc*-PR EoS proved to be impressive. Such a modern CEoS achieved an average overall deviation of less than 2% over the 306,700 available experimental data points. Excellent results are obtained for all the properties, except the molar critical volumes.
- The striking fact is that not all the NSA molecules are well modeled by the studied CEoSs and that almost two-thirds of the SA compounds of the database are represented with high accuracy by CEoSs. At this step, one might wonder whether the addition of an association term is the right solution to improve the accuracy of EoSs. The answer to this extremely complex question is not yet clear to us but we plan to publish our conclusion soon, in a next paper. At this step, it could be argued that EoSs are mostly developed and used in academia and industry to calculate phase equilibria and thermodynamic properties of fluid mixtures. This is obviously true, but we believe that this study over thousands of pure components deserves attention because, as stated by Agarwal and recalled in this article, "regardless of the sophistication of your thermodynamic model and the number of parameters in the mixing rule, you are in trouble if the vapor pressures are inaccurate. No matter how many parameters your mixture model may have, the end points of isothermal (P,x,y) phase diagrams will always be calculated incorrectly."

Finally, we hope that the proposed guidelines and database will be widely adopted by the scientific community to identify the strengths and weaknesses of a given thermodynamic model. Moreover, this database should help model developers to test new concepts in the formulation of EoSs for pure compounds.

ACKNOWLEDGMENTS

The French company TotalEnergies and more particularly Dr. Laurent Avaullée, Freddy Garcia, and Rémy Sancerry (experts in thermodynamics) are gratefully acknowledged for sponsoring this research.

AUTHOR CONTRIBUTIONS

Romain Privat Conceptualization (equal); supervision (equal); writing – review and editing (equal). **Andrés Piña-Martinez:** Software (equal); writing – original draft (equal). **Jean-Noël Jaubert:** Conceptualization (equal); supervision (equal); writing – review and editing (equal).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from DIPPR (https://www.aiche.org/dippr). Restrictions apply to the availability of these data, which were used under license for this study.

ORCID

Jean-Noël Jaubert https://orcid.org/0000-0001-7831-5684

REFERENCES

1. Jaubert J-N, Le Guennec Y, Piña-Martinez A, et al. Benchmark database containing binary-system-high-quality-certified data for cross-comparing thermodynamic models and assessing their accuracy. *Ind Eng Chem Res*. 2020;59(33):14981-15027. https://doi.org/10.1021/acs.iecr.0c01734
2. Qian J-W, Privat R, Jaubert J-N, Duchet-Suchaux P. Enthalpy and heat capacity changes on mixing: fundamental aspects and prediction by means of the PPR78 cubic equation of state. *Energy Fuel*. 2013;27(11):7150-7178. https://doi.org/10.1021/ef401605c
3. Nikolaidis IK, Privat R, Jaubert J-N, Economou IG. Assessment of the perturbed chain-statistical associating fluid theory equation of state against a benchmark database of high-quality binary-system data. *Ind Eng Chem Res*. 2021;60(24):8935-8946. https://doi.org/10.1021/acs.iecr.1c01234
4. Pina-Martinez A, Privat R, Nikolaidis IK, Economou IG, Jaubert J-N. What is the optimal activity coefficient model to be combined with the translated-consistent Peng-Robinson (tc-PR) equation of state through advanced mixing rules? Cross-comparison and grading of the Wilson, UNIQUAC and NRTL aE models against a benchmark database involving 200 binary systems. *Ind Eng Chem Res*. 2021. https://doi.org/10.1021/acs.iecr.1c03003
5. Agarwal R, Li Y-K, Satyro MA, Vieler A. Uncovering the realities of simulation. Part I. *Chem Eng Prog*. 2001;97(5):42-52.
6. Agarwal R, Li Y-K, Satyro MA, Vieler A. Uncovering the realities of simulation. Part II. *Chem Eng Prog*. 2001;97(6):64-72.
7. Soave G. Equilibrium constants from a modified Redlich-Kwong equation of state. *Chem Eng Sci*. 1972;27(6):1197-1203. https://doi.org/10.1016/0009-2509(72)80096-4
8. Peng D, Robinson D. A new two-constant equation of state. *Ind Eng Chem Fund*. 1976;15(1):59-64. https://doi.org/10.1021/i160057a011
9. Le Guennec Y, Privat R, Jaubert J-N. Development of the translated-consistent *tc*-PR and *tc*-RK cubic equations of state for a safe and accurate prediction of volumetric, energetic and saturation properties of pure compounds in the sub- and super-critical domains. *Fluid Phase Equilib*. 2016;429:301-312. https://doi.org/10.1016/j.fluid.2016.09.003
10. Piña-Martinez A, Le Guennec Y, Privat R, Jaubert J-N, Mathias PM. Analysis of the combinations of property data that are suitable for a safe estimation of consistent Twu α-function parameters: updated parameter values for the translated-consistent *tc*-PR and *tc*-RK cubic equations of state. *J Chem Eng Data*. 2018;63(10):3980-3988. https://doi.org/10.1021/acs.jced.8b00640
11. Ramírez-Vélez N, Piña-Martinez A, Jaubert J-N, Privat R. Parameterization of SAFT models: analysis of different parameter estimation strategies and application to the development of a comprehensive database of PC-SAFT molecular parameters. *J Chem Eng Data*. 2020;65(12):5920-5932. https://doi.org/10.1021/acs.jced.0c00792
12. Privat R, Jaubert J-N. Classification of global fluid-phase equilibrium behaviors in binary systems. *Chem Eng Res Des*. 2013;91(10):1807-1839. https://doi.org/10.1016/j.cherd.2013.06.026
13. Jaubert J-N, Privat R, Le Guennec Y, Coniglio L. Note on the properties altered by application of a Péneloux–type volume translation to an equation of state. *Fluid Phase Equilib*. 2016;419:88-95. https://doi.org/10.1016/j.fluid.2016.03.012
14. Privat R, Jaubert J-N, Le Guennec Y. Incorporation of a volume translation in an equation of state for fluid mixtures: which combining rule? Which effect on properties of mixing? *Fluid Phase Equilib*. 2016;427:414-420. https://doi.org/10.1016/j.fluid.2016.07.035
15. Le Guennec Y, Lasala S, Privat R, Jaubert J-N. A consistency test for α-functions of cubic equations of state. *Fluid Phase Equilib*. 2016;427:513-538. https://doi.org/10.1016/j.fluid.2016.07.026
16. Le Guennec Y, Privat R, Lasala S, Jaubert J-N. On the imperative need to use a consistent α-function for the prediction of pure-compound supercritical properties with a cubic equation of state. *Fluid Phase Equilib*. 2017;445:45-53. https://doi.org/10.1016/j.fluid.2017.04.015
17. Twu CH, Bluck D, Cunningham JR, Coon JE. A cubic equation of state with a new alpha function and a new mixing rule. *Fluid Phase Equilib*. 1991;69:33-50. https://doi.org/10.1016/0378-3812(91)90024-2

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Piña-Martinez A, Privat R, Jaubert J-N. Use of 300,000 pseudo-experimental data over 1800 pure fluids to assess the performance of four cubic equations of state: SRK, PR, *tc*-RK, and *tc*-PR. *AIChE J*. 2022;68(2):e17518. doi:10.1002/aic.17518