Fluid Phase Equilibria 612 (2027) 114820

Contents lists available at ScienceDirect

Fluid Phase Equilibria

journal homepage: www.sciencedirect.com/journal/fluid-phase-equilibria

Extension of openCOSMO-RS into a full open-source equation of state: Implementation, parameterization, and benchmarking

Jan Markgraf, D.G. Lisboa Girardi, Irina Smirnova, Simon Müller *

Institute of Thermal Separation Processes, Hamburg University of Technology, Hamburg, Germany

ARTICLE INFO

Dataset link: https://github.com/TUHH-TVT/openCOSMO-RS-Phi_cpp

Keywords:
COSMO-RS
Open-source
Parameterization
ORCA
Equation of State

ABSTRACT

The COSMO-SAC-Phi model developed by Soares et al. extends the COSMO-SAC activity-coefficient framework into a full equation of state by explicitly accounting for pressure effects. In this approach, pure substances and mixtures are represented as pseudo-mixtures consisting of the actual number of moles and an additional pseudo-component that describes free volume, or holes. In this work, we implement this extension within the openCOSMO-RS framework and evaluate it using a large and diverse set of molecules and binary systems. The resulting equation of state includes an extensive open-source parameter set with around 1800 pure-component entries, made freely available to the academic community. The four pure-component parameters were fitted to vapor-pressure and liquid-molar volume data for each substance. Model performance was assessed against two benchmark equation-of-state databases, one for pure compounds and one for binary mixtures, without introducing any binary interaction parameters. The resulting openCOSMO-RS-Phi model reproduces the accuracy of the original COSMO-SAC-Phi formulation while providing a fully open-source and accessible implementation for the scientific community. Beyond its immediate utility, it also establishes a foundation for future development of predictive EoS for electrolyte solutions.

1. Introduction

Thermodynamic models are essential for the design, development, and control of chemical processes, and the selection of an appropriate model is critical for these applications. They describe the interactions between temperature, pressure, volume, and material compositions, and are crucial to calculate thermodynamic properties and determine phase equilibrium [1]. Commonly used are equations of state (EoS), which can be applied to high pressure conditions, and activity coefficient models, also known as excess Gibbs energy (G^E) models, which perform well for liquid mixtures and are pressure independent [2].

While all thermodynamic models possess some predictive capability, Jirasek et al. [3] distinguish between two types of prediction. The first refers to frameworks that generalize across different conditions, such as temperature, pressure, and concentration, for a fixed set of components. The second refers to models that generalize over different systems. In the present work, the term 'predictive models' refers to the second category proposed by Jirasek et al. which predict properties and phase behavior of systems without requiring prior data for the specific system under consideration. This ability is particularly valuable in situations where experimental data are unavailable, scarce, difficult to obtain, or expensive.

Among the early and most widely used predictive thermodynamic models are those in which EoS are combined with G^E models. This approach merges the advantages of both EoS and G^E approaches, expanding their range of application [4,5]. This was first achieved following the publication of the Huron–Vidal mixing rule [6], which enabled the incorporation of a predictive activity coefficient model into an equation of state through a mixing rule. Since then, several mixing rules have been developed, such as MHV1, MHV2 [7], WS [8], UMR [9], UHVM [10], along with combinations of cubic EoS models like PR [11] and SRK [12] with group contribution activity coefficients such as UNIFAC [13], modified-UNIFAC [14], LIFAC [15], and ASOG [16]. This development has resulted in several useful models, including PSRK [17], LCVM [18], VTPR [19–23], UMR-PRU [24], and others [25–27]. Recently, similar approaches have been proposed for an EoS/ G^E using PC-SAFT [28–30] or CPA [31,32] as equation of state.

Other recent advancements in predictive EoS focus on the development of methods for predicting pure component parameters, binary interaction parameters, and utilizing QSPR methods for activity coefficient model [33,34]. Pure component parameter predictions are primarily carried out in SAFT EoS [35,36] and its variants, using either group contribution methods, as comprehensively reviewed by Shaahmadi and colleagues [37], or machine learning algorithms [38–43].

* Corresponding author.
E-mail address: simon.mueller@tuhh.de (S. Müller).

https://doi.org/10.1016/j.fluid.2026.114820
Received 5 May 2026; Received in revised form 29 June 2026; Accepted 16 July 2026
Available online 24 July 2026

For binary interaction parameters, predictive EoS have been developed for PR EoS using group contribution models, leading to models like PPR78 [44] and EPPR78 [45,46], as well as QSPR [47] and machine learning [48]. Similarly, both QSPR [49] and machine learning [50] approaches have been applied to PC-SAFT.

The conductor-like screening model for real solvents (COSMO-RS) [51–53] is a completely a priori predictive activity coefficient model. It treats a molecule as a collection of surface segments and calculates the molecular chemical potential by summing the chemical potentials of each segment. These segment potentials are determined by the interaction energies between segment screening charges, obtained through quantum chemical *ab initio* solvation calculations. In contrast to other established group contribution G^E models, such as UNIFAC and modified-UNIFAC, COSMO-RS requires a much smaller set of parameters and can partially differentiate between isomers [54]. Several versions of the COSMO-RS model have emerged in the literature, such as COSMO-SAC [55], COSMO-RS(Ol) [54], COSMO-vac [56], COSMO-SAC-dsp [57] and the closely related F-SAC [58,59].

Although COSMO-based models are effective, they share the same limitations as other G^E models, as they do not account for pressure effects and are restricted to representing incompressible liquids. As a result, various approaches have been developed to integrate them with existing EoS. Early methods involved using an EoS/ G^E framework with UNIQUAC as the activity coefficient model, with parameters adjusted to either COSMO-RS or COSMO-SAC [60,61]. Within the EoS/ G^E framework, recent studies have explored coupling cubic EoS with COSMO-based models through both traditional and novel mixing rules [62–72], and PC-SAFT with COSMO-RS has been implemented as well [73]. Other significant contributions include NRCOSMO [33,74,75], based on the Non-Random Hydrogen-Bonding EoS; σ-MTC [76,77], derived from the Mattedi–Tavares–Castier EoS; COSMO-SAC-Phi [78], a new equation of state inspired by lattice-fluid theory; and the recent notes on the essentials for the formulation of a COSMO-RS-based equation of state published by Panayiotou and colleagues [79].

In this work, an open source implementation of the original COSMO-SAC-Phi is proposed as an extension of the openCOSMO-RS [80] activity coefficient model. COSMO-SAC-Phi is a unique and novel approach that, unlike most of the literature, expands the COSMO-SAC model into a full equation of state rather than coupling it with or modifying existing EoS. Consequently, the resulting EoS does not rely on binary interaction parameters to describe the mixtures, but uses the underlying COSMO-RS theory. Although an improved version of COSMO-SAC-Phi exists [81], featuring an additional parameter to increase precision and thermodynamic consistency, it was not employed in this study. Another related model is F-SAC-Phi [82] which uses the F-SAC model instead of the COSMO-SAC model as backend. To prevent any ambiguity when referring to the original or to the open-source model, the proposed method is designated as openCOSMO-RS-Phi. This work serves as a foundation for the further development of the first a priori predictive equation of state based on COSMO-RS theory for mixtures.

2. Methodology

This work follows a three-step approach. First, COSMO-SAC-Phi is implemented in the open-source form of openCOSMO-RS-Phi. To this end, the theoretical framework is revisited in this section, and openCOSMO-RS-Phi is formulated as a Helmholtz-explicit equation of state. For the assessment of model performance, pure-component parameters are fitted, and the resulting model is benchmarked against both a pure-component database and a binary-mixture database. The corresponding methodological and theoretical foundations are presented in the second part of this section.

2.1. Mathematical derivation

openCOSMO-RS-Phi, as an open-source implementation of COSMO-SAC-Phi, is a perturbation model designed to predict the behavior of fluids. It employs an attraction term developed by integrating COSMO-RS with concepts from lattice-fluid theory. This concept introduces holes that occupy all available free volume and account for pressure and volume effects. Therefore, all systems are considered to be pseudo-mixtures from the moles of real mixture or pure substance $\mathbf{n} = [n_1, n_2, \dots, n_i, \dots, n_N]$, and the moles of holes n_h, with molar vector $\tilde{\mathbf{n}} = [n_1, n_2, \dots, n_i, \dots, n_N, n_h]$. This leads to the following expression for the total system volume V, which, given the total moles of the real fluid, enables the calculation of the moles of holes:

$$V = \sum_i n_i b_i + n_h b_h \tag{1}$$

$$n_h = \frac{V - \sum_i n_i b_i}{b_h} \tag{2}$$

where b_i is the molar cavity of species i and b_h is the molar cavity of a hole. The values from b_i and b_h are the first among the four existing parameters that need to be adjusted.

Although the concept of vacancy/holes within COSMO-SAC was initially introduced with the COSMO-vac model [56], a decade prior to the development of COSMO-SAC-Phi, it is applied only to computing the activity coefficients of solutes diluted in a supercritical phase. Rather than describing the vacancy as an additional pseudo-substance, COSMO-vac calculates a vacancy fraction α_{vac} that modifies the COSMOSPACE equation described at Eq. (29). Another similar approach is the use of a vacuum pseudo-liquid within the COSMOtherm software [83]. In the latter case, the vacuum molecule is used only to approximate the gas phase at a liquid–gas interface. Similarly to COSMO-SAC-Phi, this pseudo-substance is characterized by the presence of a single surface segment with a fixed area, and the possession of volume.

It is possible to derive the relevant residual properties from the residual Helmholtz energy $A^r(T, V, \mathbf{n})$. The residual Helmholtz energy $A^r(T, V, \mathbf{n})$ is defined as the difference between the Helmholtz energy of the real phase and the Helmholtz energy of an ideal gas mixture having the same temperature T, the same volume V, and the same composition as the real mixture. The openCOSMO-RS-Phi equation for A^r can be divided into a repulsive contribution, A^r_R, and an attractive contribution, A^r_A:

$$A^r(T, V, \mathbf{n}) = A^r_R(T, V, \mathbf{n}) + A^r_A(T, V, \mathbf{n}) \tag{3}$$

The repulsive interactions are represented using the classical Van der Waals approach. This simplified approximation yields [84]:

$$P_R = \frac{nRT}{V - \sum_i n_i b_i} \tag{4}$$

It is important to note that the repulsive pressure P_R according to Eq. (4) is actually the sum of the repulsive contribution to the residual pressure P^r_R and the pressure of an ideal gas mixture P^{IGM}: $P_R = P^r_R + P^{IGM}$. A^r_R can then be calculated from classical thermodynamics as follows:

$$A^r_R(T, V, \mathbf{n}) = -\int_\infty^V \frac{nRT}{V - \sum_i n_i b_i} - \frac{nRT}{V} \mathrm{d}V \tag{5}$$

$$= -nRT \ln\left(\frac{V - \sum_i n_i b_i}{V}\right) \tag{6}$$

The attractive contribution, A^r_A, can be obtained from Euler's theorem, provided that the Helmholtz energy is a homogeneous function to the first degree in V and n which is correct within the frame of openCOSMO-RS-Phi:

$$A^r_A(T, V, \mathbf{n}) = \left(\frac{\partial A^r_A}{\partial V}\right)_{T,\mathbf{n}} V + \sum_{i=1}^{N} n_i \left(\frac{\partial A^r_A}{\partial n_i}\right)_{T,V,n_j} \tag{7}$$

An important detail when converting the partial derivative of the attractive contribution to the Helmholtz energy of the real fluid A^r_A into the partial derivative of the Helmholtz energy of the pseudo-mixture $\tilde{A}^r_A$ in regards to the moles of holes, is that free volume changes are governed completely in the pseudo-mixture by the variation of

the amount of holes, while the overall volume of the pseudo-mixture remains constant. Therefore:

$$\left(\frac{\partial A_A^r}{\partial V}\right)_{T,\mathbf{n}} = \left(\frac{\partial \tilde{A}_A^r}{\partial n_h}\right)_{T,V,\mathbf{n}} \left(\frac{\partial n_h}{\partial V}\right)_{T,\mathbf{n}} \tag{8}$$

The attractive residual chemical potential of the holes in the pseudo-mixture is defined:

$$\tilde{\mu}_{Ah}^r = \left(\frac{\partial \tilde{A}_A^r}{\partial n_h}\right)_{T,V,\mathbf{n}} \tag{9}$$

Moreover, differentiating Eq. (2) in regards to the volume while **n** is constant yields the final term of Eq. (8):

$$\left(\frac{\partial n_h}{\partial V}\right)_{T,\mathbf{n}} = \frac{1}{b_h} \tag{10}$$

Thus, and considering for perturbation terms, pressure due to attractive forces is derived from the fundamental thermodynamic relations:

$$\left(\frac{\partial A_A^r}{\partial V}\right)_{T,\mathbf{n}} = -P_A = \frac{\tilde{\mu}_{Ah}^r}{b_h} \tag{11}$$

In Eq. (7), the partial derivative of the attractive Helmholtz contribution with respect to the amount of species i must be reformulated for the pseudo-mixture. In this framework, a change in n_i affects the attractive contribution both directly through species i and indirectly through the number of holes. Since the total volume of the pseudo-mixture is kept constant, any change in the number of holes is solely induced by the change in n_i:

$$\left(\frac{\partial A_A^r}{\partial n_i}\right)_{T,V,n_j} = \left(\frac{\partial \tilde{A}_A^r}{\partial n_i}\right)_{T,V,n_j,n_h} + \left(\frac{\partial \tilde{A}_A^r}{\partial n_h}\right)_{T,V,\mathbf{n}} \left(\frac{\partial n_h}{\partial n_i}\right)_{T,V,n_j} \tag{12}$$

The attractive residual chemical potential of species i in the pseudo-mixture is defined:

$$\tilde{\mu}_{Ai}^r = \left(\frac{\partial \tilde{A}_A^r}{\partial n_i}\right)_{T,V,n_j,n_h} \tag{13}$$

Although the hole molar cavity b_h has thus far been treated as constant, it is represented by a sum of the contribution of the hole molar volume of each molecule $b_{h,i}$ within the system through the following mixing rule:

$$b_h = \sum_i \frac{n_i}{n} b_{h,i} \tag{14}$$

Taking this into consideration, the final term of Eq. (12) can be obtained by differentiating Eq. (2) in regards to the n_i while V and n_j are constant:

$$\left(\frac{\partial n_h}{\partial n_i}\right)_{T,V,n_j} = -\left[\frac{b_i}{b_h} + \frac{n_h}{n}\left(\frac{b_{h,i}}{b_h} - 1\right)\right] \tag{15}$$

Therefore, the chemical potential of the components in the real mixture are calculated from the chemical potentials of the pseudo-mixture as follows:

$$\left(\frac{\partial A_A^r}{\partial n_i}\right)_{T,V,n_j} = \tilde{\mu}_{Ai}^r - \tilde{\mu}_{Ah}^r \left[\frac{b_i}{b_h} + \frac{n_h}{n}\left(\frac{b_{h,i}}{b_h} - 1\right)\right] \tag{16}$$

If $b_{h,i}$ have the same value throughout all the components of the mixture, this is to say $b_{h,i} = b_h$, Eq. (16) can be simplified to:

$$\left(\frac{\partial A_A^r}{\partial n_i}\right)_{T,V,n_j} = \tilde{\mu}_{Ai}^r - \tilde{\mu}_{Ah}^r \frac{b_i}{b_h} \tag{17}$$

For both pure components and mixtures, Eq. (7) becomes:

$$A_A^r(T,V,\mathbf{n}) = \sum_i n_i \left(\tilde{\mu}_i^r - \frac{b_i}{b_h}\tilde{\mu}_h^r\right) - P_A V \tag{18}$$

Combining Eqs. (6) and (18) yields the complete residual Helmholtz energy, making openCOSMO-RS-Phi a Helmholtz-explicit EoS. As a consequence of linear homogeneity of A^r, one directly obtains a relation for the molar residual Helmholtz energy a^r as a function of temperature T, molar volume v and mole fractions x_i:

$$a^r(T,v,\mathbf{x}) = \sum_i x_i \left(\tilde{\mu}_i^r - \frac{b_i}{b_h}\tilde{\mu}_h^r\right) - P_A v - RT \ln\left(\frac{v - \sum_i x_i b_i}{v}\right) \tag{19}$$

At a defined temperature T, the pressure P is described as the sum of the reference fluid repulsive term and the perturbation attractive term [85]:

$$P = P_R + P_A \tag{20}$$

where P_R is the repulsive contribution pressure, described in Eq. (4), and P_A is the attractive contribution pressure, shown in Eq. (11). Therefore:

$$P = \frac{nRT}{V - \sum_i n_i b_i} - \frac{\tilde{\mu}_{Ah}^r}{b_h} \tag{21}$$

At the same temperature, the fugacity coefficient ϕ_i of species i can be described by [86]:

$$\ln \phi_i = \frac{1}{RT}\left(\frac{\partial A^r}{\partial n_i}\right)_{T,V,n_j} - \ln Z \tag{22}$$

where A^r is the residual Helmholtz energy, $Z = PV/nRT$ is the compressibility factor, and $n = \sum_i n_i$ is the molar amount of the real fluid.

Similarly to pressure, the partial derivative of residual Helmholtz energy in regards to the moles of species i can be described as the sum of a repulsive and an attractive contributions [85]:

$$\left(\frac{\partial A^r}{\partial n_i}\right)_{T,V,n_j} = \left(\frac{\partial A_R^r}{\partial n_i}\right)_{T,V,n_j} + \left(\frac{\partial A_A^r}{\partial n_i}\right)_{T,V,n_j} \tag{23}$$

where:

$$\left(\frac{\partial A_R^r}{\partial n_i}\right)_{T,V,n_j} = RT\left[\ln\frac{V}{V - \sum_i n_i b_i} + \frac{nb_i}{V - \sum_i n_i b_i}\right] \tag{24}$$

Using Eqs. (24) and (16), the fugacity coefficient of component i in the real mixture is obtained through the following expressions:

$$\ln \phi_i(T,V,\mathbf{n}) = \tilde{\mu}_{Ai}^r - \tilde{\mu}_{Ah}^r\left[\frac{b_i}{b_h} + \frac{n_h}{n}\left(\frac{b_{h,i}}{b_h} - 1\right)\right] + \ln\frac{V}{V - \sum_i n_i b_i} + \frac{nb_i}{V - \sum_i n_i b_i} - \ln\frac{PV}{nRT} \tag{25}$$

The expressions for P_A, P_R, and $\ln\phi_i$ relevant to the application of the EoS have so far been written in terms of T, V, and n_j. However, all of these expressions are homogeneous of the order zero with respect to the number of moles and may then also be expressed in terms of T, the molar volume v, and the mole fractions $x_j = n_j/n$.

Finally, to completely describe attractive terms, the residual chemical potential in the pseudo-mixture of species i and of the holes are similarly defined. To put it differently, the mathematical foundation starting in Eq. (26) is applicable to both species i and holes. However, for the purpose of clarity, their derivation will be presented solely for species i, with brief comments provided when the derivation for holes differ. Therefore, the attractive chemical potential of species i is related to the activity a_i by:

$$\tilde{\mu}_{Ai} = \tilde{\mu}_{Ai}^* + RT \ln a_i \tag{26}$$

where $\tilde{\mu}_{Ai}^*$ is the attractive chemical potential of species i in a defined reference state. In the model, the reference state used is the ideal gas, thus:

$$\tilde{\mu}_{Ai}^r = RT \ln a_i \tag{27}$$

The activity of species i for COSMO-RS is defined as:

$$\ln a_i = \sum_I \left[\frac{\mathcal{A}_I^i}{a_{eff}} \ln\left(\frac{\Gamma_I}{\Gamma_I^*}\right)\right] \tag{28}$$

where $\mathcal{A}_I^i$ is the surface area of segment I of molecule i, a_{eff} is the average molecular contact area Γ_I is the activity coefficient of segment I, and Γ_I^* is the activity coefficient of segment I at the reference state. The area of one hole is kept equal to that of a 1 Å radius sphere. Thus, $\mathcal{A}_I^h = 12.57\ \text{Å}^2$.

Through the COSMOSPACE equation system [87], the segment activity coefficient can be calculated:

$$\Gamma_I = \frac{1}{\sum_J X_J \Gamma_J \tau_{IJ}} \tag{29}$$

where X_J is the fraction of segment J that can be found within all segments, and τ_{IJ} is the interaction parameter from segments I and J. They are defined as:

$$X_J = \frac{\sum_i x_i \mathcal{A}_J^i}{\sum_i \sum_I x_i \mathcal{A}_I^i} \tag{30}$$

$$\tau_{IJ} = \exp\left(-\frac{G_{IJ}^{int}}{RT}\right) \tag{31}$$

where $\mathcal{A}_J^i$ is the surface area of segment J of molecule i, x_i is the mole fraction of molecule i, and G_{IJ}^{int} describes the interaction energy between segments J and I. For openCOSMO-RS-Phi, the mole fraction of molecule i in the pseudo-mixture, $\tilde{x}_i = n_i/(n+n_h)$, is used in Eq. (30).

In this framework, where ideal gas is considered as the reference state, Γ_I^* is calculated at infinite molar volume. Hence, in ideal gas, the mole fractions of all species are considered to approach the limit zero, while the mole fraction of the holes is regarded to approach the limit of one:

$$\Gamma_I^* = \Gamma_I^{IG} \implies \left(\forall i \in \{1, 2, \dots, n\},\ x_i = 0\right) \wedge \left(x_h = 1\right) \tag{32}$$

Furthermore, the interaction energy between segments remains to be described. This step marks the primary distinctions between COSMO-SAC-Phi and openCOSMO-RS-Phi. Both models describe the interaction energy as a sum of the contribution from electrostatic misfit interactions E_{IJ}^{mf}, the contribution from hydrogen bond interactions G_{IJ}^{hb}, and the contribution from dispersion forces G_{IJ}^{Disp}, although the formulation for these individual free energies differ:

$$G_{IJ}^{int} = E_{IJ}^{mf} + G_{IJ}^{hb} + G_{IJ}^{Disp} \tag{33}$$

Among the three contributions, both the dispersion forces energy and the electrostatic misfit interactions energy are represented identically as the original model. The first is described as a simple combining rule from two distinct segment dispersions from species i, δ_I^i and δ_J^i, which are temperature dependent:

$$G_{IJ}^{Disp} = -\frac{1}{2}\sqrt{\delta_I^i \delta_J^i} \tag{34}$$

$$\delta_I^i = \delta^{i,0}\left[1 - \exp\left(-\frac{\delta^{i,T}}{T}\right)\right] \tag{35}$$

where $\delta^{i,0}$ and $\delta^{i,T}$ are parameters without assigned names, they are adjusted for each individual component, rather than to each segment. As dispersion is not considered for holes, δ^h is equal to zero.

The misfit contribution is calculated from the charge densities of segments I and J, denoted as σ_I and σ_J:

$$E_{IJ}^{mf} = \frac{a_{eff}\alpha}{2}\left(\sigma_I + \sigma_J\right)^2 \tag{36}$$

with α as the misfit prefactor.

With regard to the calculation of the hydrogen bond interaction energy, the approach adopted is analogous to that proposed by Gerlach, et al. [80] in the derivation of openCOSMO-RS. However, it differs in that the hydrogen bond strength parameter c_{hb} is considered temperature-independent. This change was made to improve the numerical performance and bring the implementation closer to the original publication of the model.

$$G_{IJ}^{hb} = \begin{cases} c_{hb} a_{eff}\left[\min\left(0; \sigma_I + \sigma_{hb}\right) \cdot \max\left(0; \sigma_J + \sigma_{hb}\right)\right] & \text{if } \sigma_I < \sigma_J \\ c_{hb} a_{eff}\left[\min\left(0; \sigma_J + \sigma_{hb}\right) \cdot \max\left(0; \sigma_I + \sigma_{hb}\right)\right] & \text{if } \sigma_I \geq \sigma_J \end{cases} \tag{37}$$

Table 1
Universal constants of openCOSMO-RS.

Parameter	Unit	Value
a_{eff}	AA^2	4.857
α	$\text{kJ} \cdot \text{Å}^2/(\text{mol} \cdot \text{e}^2)$	8130
c_{hb}	$\text{kJ} \cdot \text{Å}^2/(\text{mol} \cdot \text{e}^2)$	42 564
σ_{hb}	$\text{e}/\text{Å}^2$	9.443×10^{-3}

where c_{hb} is the hydrogen bond strength parameter, and σ_{hb} is hydrogen bond threshold parameter.

For a more thorough overview of the openCOSMO-RS model and its derivation, please refer to the original publication [80]. Its open-source workflow is maintained in this work.

2.2. Parameterization of the openCOSMO-RS model

Table 1 lists the optimized global parameters used in this study for the underlying openCOSMO-RS model. In contrast to previous parameterizations [80,88], c_{hb}^{T} was set to 0 and the basic misfit term was used instead of the improved one.

2.3. Experimental databases and optimization for pure compound input

There are four parameters that need to be defined for each pure component: the molar cavity of the component b_i, the hole molar volume of said component $b_{h,i}$, and the two dispersion parameters $\delta^{i,0}$ and $\delta^{i,T}$. They are obtained from the minimization of the following objective function using experimental vapor pressure P_i^{sat} and molar volume v_i^l:

$$OF = \frac{1}{N_p}\sum_i w_P\left(\frac{P_i^{sat} - P_i^{sat,calc}}{P_i^{sat}}\right)^2 + w_v\left(\frac{v_i^l - v_i^{l,calc}}{v_i^l}\right)^2 \tag{38}$$

Here, N_p is the number of experimental points, calculated entries are indicated by the *calc* superscript, and w_P and w_v are the weighting factors for the saturated pressure and molar volume, respectively.

The σ-profile (screened molecular surface) is generated by the quantum-chemical calculation in ORCA, in which the molecular cavity is discretized into individual surface segments. The total molecular surface area is calculated as the sum of the areas of these segments. These quantum-chemically calculated surface quantities are retained in openCOSMO-RS and are independent of the adjustable molar cavity parameter used in the equation of state. For more details the reader is referred to the original publication of the free and open source conformer pipeline [88].

Experimental data for the vapor pressure (P^{sat}) and molar volumes (v^{liq}) of pure components were collected from the DIPPR 801 database [89]. The guidelines proposed by Piña-Martinez et al. [90] were applied over the database, limiting it to 1800 pure fluids. Furthermore, ten additional substances were removed from the analysis, either because they were mixtures (air), salts or organometallic (ammonium chloride, ammonium sulfide, iron pentacarbonyl, mercury dichloride, nickel carbonyl, vanadium oxytrichloride, vanadium tetrachloride), or lead to convergence issues (eicosamethylnonasiloxane, hexacosamethyldodesiloxane). Similarly to COSMO-SAC-Phi, the present study has focused exclusively on data constrained within a specified temperature range. In this instance, the analysis has been conducted within the temperature range of 50 % to 90 % of the critical temperature, with the data evaluated at 10 K intervals. Thus, for each component, vapor pressure and liquid molar volume are evaluated on the same temperature grid.

The performance of openCOSMO-RS-Phi for pure components was studied by calculating the mean absolute percentage error (MAPE) for each property x, represented by Eq. (39). This method was previously

used for COSMO-SAC-Phi, as well as in the aforementioned work by Piña-Martinez et al. [90].

$$\mathrm{MAPE}_x = \frac{100}{N_p}\sum_i \left|\frac{x_i - x_i^{calc}}{x_i}\right| \tag{39}$$

Piña-Martinez et al. propose a classification of 'well-modeled' and 'badly-modeled' based on the distribution of the calculated data, analyzing over the complete dataset, as well over the self-associating (SA) and non-self-associating (NSA) fluids. In Eq. (40), a component i is considered 'well-modeled' if both conditions are satisfied; σ is the standard deviation and the NSA superscript indicates that only non-self-associating components are considered. Despite its limitations, which stem from its reliance on the distribution of the reference set rather than absolute accuracy and its omission of an evaluation of meaningful precision requirements, the analysis was conducted. To circumvent these concerns, the results are also categorized into four performance levels based on the MAPE for both vapor pressure and molar volume: 'excellent', 'good', 'satisfactory' and 'poor'. A model is classified as 'excellent' when the MAPE for both properties is ≤1 %, 'good' when ≤5 %, and 'satisfactory' when ≤10 %. Cases in which the MAPE for either property exceeds 10 % are considered 'poor'. The coefficient of determination, R^2, is also available.

$$\begin{cases} \left(\mathrm{MAPE}_{P^{sat}}\right)_i \le \overline{\mathrm{MAPE}}^{\mathrm{NSA}}_{P_{sat}} + \sigma^{\mathrm{NSA}}_{P_{sat}} \\ \left(\mathrm{MAPE}_{v^{liq}}\right)_i \le \overline{\mathrm{MAPE}}^{\mathrm{NSA}}_{v^{liq}} + \sigma^{\mathrm{NSA}}_{v^{liq}} \end{cases} \tag{40}$$

Initially, the weighting factors were set to be the same as those used by Soares et al. [78]. However, analysis of a larger group of molecules revealed bias towards low MAPE values for vapor pressure, with high MAPE for molar volumes. An alternative set of weighting factors proposed in Ramírez-Vélez et al. [91] evaluation of the same 1800 pure fluids was examined. Nevertheless, given the significant impact of the weighting factors on the average MAPE of each property, and in the value of the objective function, a new approach is proposed in order to obtain less arbitrary weights. In this approach, the MAPEs of each property for 4 different EoS (tc-RK, tc-PR, PC-SAFT, and I-PC-SAFT) [90,91] using the same component data were used. This strategy maintains the bias towards saturated pressure present in the aforementioned works while ensuring that molar volume data are not underrepresented. The weighting factors are then defined at Eq. (41) as a normalized inverse of the average MAPE of the respective property X. Their values are roughly 0.67 for w_P and 0.33 for w_v.

$$w_X = \frac{\frac{1}{\overline{\mathrm{MAPE}}_X}}{\frac{1}{\overline{\mathrm{MAPE}}_{P^{sat}}} + \frac{1}{\overline{\mathrm{MAPE}}_{v^{liq}}}} \tag{41}$$

The identification of the optimal set of parameters was achieved through the implementation of two minimization methods, executed in a three-step process. Initially, the Nelder–Mead method [92] was implemented, with the parameters constrained to values that were more restrictive and closer to the results presented by Soares et al. [78]. The initial values for both b_i and b_h were respectively obtained through linear and quadratic regression between the parameter values and the COSMO volume, using the data from the COSMO-SAC-Phi publication. The starting values for both $\delta^{i,T}$ and $\delta^{i,0}$ were obtained by calculating the mean value of each parameter from the same publication. The best parameter set obtained from the initial step was used as starting values for the subsequent step, which incorporated Nelder–Mead minimization once more, albeit without bounds. This was done intentionally to circumvent the occurrence of local minima in proximity to the boundaries, thereby establishing a suitable initial point for an broader parameter evaluation. The final minimization step was executed via differential evolution [93] with a range of bounds for each parameter that was wider than those from the first step. For each substance, the parameter set with the smallest value from the objective function of the three described steps was selected. The bounds for each adjusted parameter in the first and third steps are shown in Table 2, with the COSMO volume being represented as v_{COSMO}.

2.4. Benchmarking database from Jaubert et al. [94]

The assessment of openCOSMO-RS-Phi for binary mixtures is based on a high-quality reference database proposed by Jaubert and coworkers [94]. To cover a broad range of system types, the authors selected 107 nonelectrolytic pure compounds, forming a total of 200 binary mixtures. Its classification by association behavior and standardized evaluation procedure allow for a rigorous, systematic comparison of thermodynamic models over a broad range of mixtures, ensuring that predictive performance is assessed on physically meaningful and practically relevant systems. Pure components are classified according to their association behavior (where "association" refers exclusively to hydrogen bonding) as follows:

- Non-Associating (NA): neither a hydrogen atom capable of hydrogen bonding nor a lone pair available.
- Hydrogen-Donor (HD): a hydrogen atom capable of hydrogen bonding available but no lone pair of electrons.
- Hydrogen-Acceptor (HA): lone pair of electrons available but no hydrogen atom capable of hydrogen bonding.
- Self-Associating (SA): hydrogen atom capable of hydrogen bonding and a lone pair of electrons available.

Based on this classification, binary systems are divided into nine groups, each assigned a binary association code (BAC), as summarized in Table S1.

The dataset of the 200 binary systems contains low- and high-pressure phase-equilibrium data (VLE, LLE, VLLE, azeotropic, and critical points) as well as energetic data, namely enthalpy (h^{M}) and heat-capacity ($c_{\mathrm{p}}^{\mathrm{M}}$) changes upon mixing.

To quantify a model's accuracy against this database, the original publication also introduced a standard evaluation procedure. Subsequently, two modifications to this grading procedure were proposed by Piña-Martinez et al. [95]. The first modification concerns the grading of three-phase-line data. The second introduces a Success Ratio (SR) into the grading scheme in order to account for out-of-model (OM) data points. OM points are points for which a MAPE cannot be evaluated because of an incorrect qualitative prediction.

For the exact methodology, the reader is referred to the original publications [94,95] or the Supporting Information. In brief:

1. For each BAC, the MAPE is determined for each property type (10 at the most).
2. The MAPE is then converted into a mark ranging from 0 and 20 for each BAC, with higher values indicating better model performance.
3. After calculating the scores for each BAC and each property, three averaging steps are performed to obtain the final model score. First, a mark is assigned to each BAC by averaging the up to ten property scores for that BAC. By averaging the scores of some BACs, a mark is then calculated for each of in total four categories: non-associating (N-A: BAC_1–BAC_4), self-associating (S-A: BAC_5), cross-associating (C-A: BAC_6), and combined self- and cross-associating systems (S-A + C-A: BAC_7–BAC_9). Finally, the overall mark of the model is obtained as the average of the four category marks.

The final mark of a thermodynamic model quantifies its predictive capability and is already available for several EoS, including the translated-consistent Peng–Robinson equation of state combined with different activity-coefficient models through advanced mixing rules [90,96], as well as PC-SAFT with and without binary interaction parameters (BIPs) [97,98].

Table 2
openCOSMO-RS-Phi adjusted parameters boundaries for minimization methods.

Adjusted	Nelder–Mead		Differential evolution	
parameters	min	max	min	max
b_i	0.5 v_{COSMO}	2 v_{COSMO}	0.5 v_{COSMO}	2 v_{COSMO}
$b_{h,i}$	2	0.1 v_{COSMO} + 20	8	0.1 v_{COSMO} + 20
$\delta^{i,0}$	0.05	6	0.05	0.005 v_{COSMO} + 8
$\delta^{i,T}$	0.05	30	0.05	30

In the present work, openCOSMO-RS-Phi is evaluated using zero BIPs and the modified grading procedure proposed by Piña-Martinez et al.; the model is assessed based on the resulting marks. It should be emphasized that, since no BIPs were regressed, the results are true predictions.

2.5. Thermodynamic calculations

This section briefly explains how the 10 properties of the Jaubert database are calculated.

Two-phase data. VLE calculations are done by first generating the complete VLE region of the binary phase diagram using either an isothermal or isobaric, fast-converging bubble point algorithm consisting of two iterative loops [99]. This approach offers two main advantages. First, so-called *out of model* points of type 1 according to Jaubert et al. [94], i.e. a two-phase system is experimentally observed but model predicts one single phase, can be readily identified. Second, the procedure provides good initial estimates for the K values, which can subsequently be used in a PT-flash calculation to determine the vapor- and liquid-phase compositions at a specified experimental pressure and temperature.

A necessary condition for the occurrence of LLE is the existence of a local maximum in the plots of the activity of component 1, a_1, and the activity of component 2, a_2, versus mole fraction. Therefore, the $a_i(x_1)$ curves were first calculated using openCOSMO-RS-Phi. If both curves exhibit a local maximum, an LLE PT-flash is carried out using the initialization scheme proposed by Ohanomah and Thompson [100].

Three-phase line. When only the temperature is specified, the three-phase pressure and composition is calculated using the horizontal line method described by Elliott and Lira [99]. In this approach, a conventional bubble-point pressure algorithm for VLE is first applied. Odd loops and self-intersection of the dew curve that might then be observed are characteristic of a VLLE. A horizontal line through the intersection point of the dew curve then gives the three-phase line data. The complete binary phase diagram is obtained by overlapping the VLE diagram and the LLE diagram. If the method outlined above does not work, this overlap also provides an alternative means of determining the three-phase line, as the VLLE occurs at that pressure where the LLE curve intersects the bubble-point curve.

Critical and azeotropic points. The calculation of the critical and azeotropic points is performed by systematically tracing the coexistence curves in the isothermal phase diagram until the difference in the mole fraction of component 1 between the two coexisting phases is smaller than 0.00015.

Mixing functions. As detailed in [101,102], the mixing function of any intensive thermodynamic property is equal to the difference between the actual property d of the mixture (for $P, T, \mathbf{z}$) and the mole-fraction-weighted sum of the properties of the pure components considered in their stable state at equal P and T. It is therefore necessary to determine the aggregation state of the mixture and the stable state of the pure components. To determine the phase state of the mixture, the bubble- and dew-point pressures at the given mixture composition $\mathbf{z}$ and temperature T are calculated. If $P_{\text{dew}} \leq P \leq P_{\text{bubble}}$, a subsequent flash calculation is performed to obtain the compositions of the two phases. If $P \geq P_{\text{bubble}}$, the mixture exists entirely as a compressed (subcooled) liquid. In this liquid case, an additional check for liquid–liquid equilibrium (LLE) is performed. If $P \leq P_{\text{dew}}$, the mixture exists entirely as superheated vapor.

The identification of the stable state for the pure substance at the same T, P is carried out in a similar way by comparing the pressure P with the vapor pressure of the pure component at temperature T.

In the most general (i.e. multi-phase fluid) case, the mixing property d^{M} (here $d \in \{h, c_{\text{p}}\}$) at fixed temperature, pressure, and composition is obtained from a linear combination of the individual phases' mixing functions using the molar proportions θ_k of phase k as weighting factors:

$$d^{\text{M}} = \sum_k \theta_k \cdot d_k^{\text{M}}(T, P, \mathbf{x}^{(k)}) \tag{42}$$

In Eq. (42), $\mathbf{x}^{(k)}$ denotes the composition vector of phase k.

According to Qian et al. [102], a mixing function at constant pressure is connected to the residual properties as follows:

$$d^{\text{M}}(T, P, \mathbf{z}) = d^{\text{r}}(T, v, \mathbf{z}) - \sum_i^N z_i \cdot d_{\text{pure } i}^{\text{r,stable state}}(T, v_{\text{pure } i}) \tag{43}$$

where $v_{\text{pure } i}$ is the molar volume of pure compound i at the same temperature T and the same pressure P as the mixture.

Residual molar enthalpy, h^{R}, can be calculated by using the following thermodynamic relation:

$$h^{\text{r}}(T, v, \mathbf{z}) = a^{\text{r}}(T, v, \mathbf{z}) - T\left(\frac{\partial a^{\text{r}}}{\partial T}\right)_{v,\mathbf{z}} + P(T, v, \mathbf{z})v - RT \tag{44}$$

The residual heat capacity at constant pressure, c_{p}^{r}, is given by the following expression:

$$c_{\text{p}}^{\text{r}}(T, P, \mathbf{z}) = -T\left(\frac{\partial^2 a^{\text{r}}}{\partial T^2}\right)_{v,\mathbf{z}} - R - T\left(\frac{\left[\left(\frac{\partial P}{\partial T}\right)_{v,\mathbf{z}}\right]^2}{\left(\frac{\partial P}{\partial v}\right)_{T,\mathbf{z}}}\right) \tag{45}$$

As can be seen from Eqs. (43) and (44), h^{r} and c_{p}^{r} are derivative properties and the derivates with respect to temperature present in these equations include terms accounting the temperature dependence of the segment activity coefficients Γ. It follows that both the first and second partial derivative of the COSMOSPACE equation with respect to temperature are in $\left(\frac{\partial \Gamma}{\partial T}\right)_{v,\mathbf{z}}$, or $\left(\frac{\partial^2 \Gamma}{\partial T^2}\right)_{v,\mathbf{z}}$, respectively and the exact same linear equation system as obtained by Yan [103] through an optimization approach has to be solved. It is worth emphasizing that all derivatives in Eqs. (43) and (44) can be calculated analytically. In summary, an accurate estimation of mixing properties requires good description of both the two-phase split (i.e., the phase diagram) and the temperature dependence of fugacity coefficients and phase equilibrium.

3. Results and discussions

3.1. Pure components

3.1.1. Comparison with COSMO-SAC-Phi

The performance is first assessed against the original COSMO-SAC-Phi publication. As shown in Fig. 1, comparison for the 40 components reported in the original publication indicates that COSMO-SAC-Phi and openCOSMO-RS-Phi yield similar results for vapor-pressure prediction,

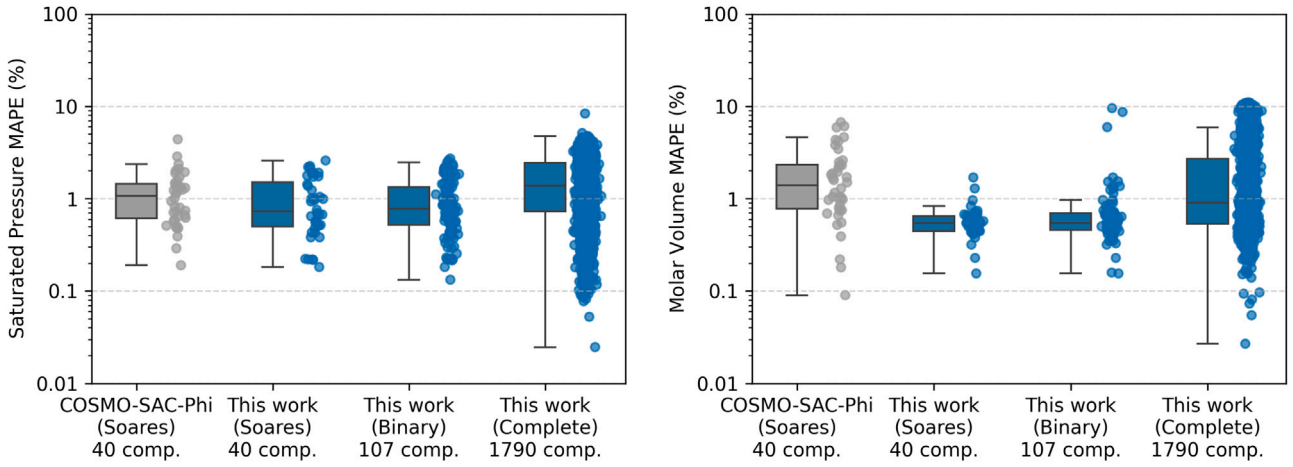

Fig. 1. Comparison of the resulting MAPE of vapor pressure and molar volume from openCOSMO-RS-Phi across different sets of components and COSMO-SAC-Phi.

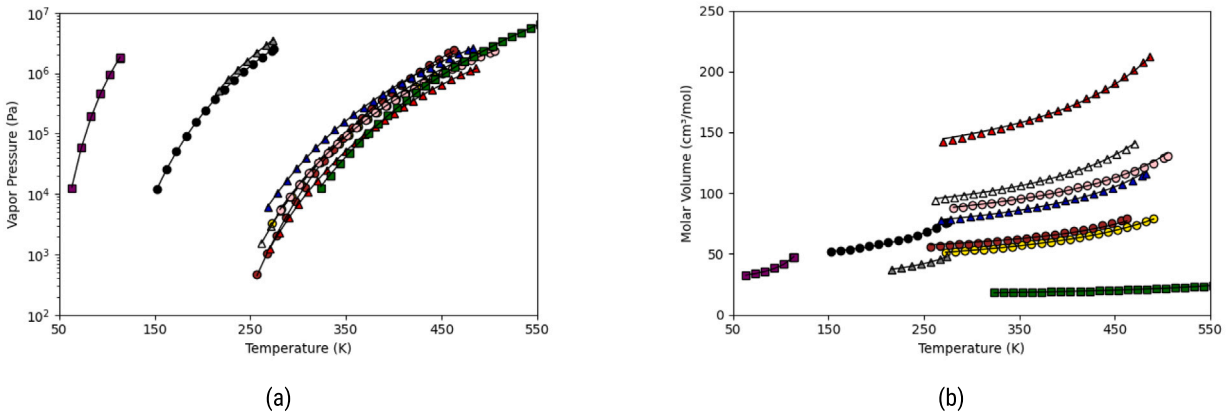

Fig. 2. Vapor pressure (a) and molar volume (b) for 10 exemplary components.

whereas the latter provides a substantial improvement in molar-volume calculations. Three molecular datasets are presented in this figure: 'Soares' refers to the set of molecules for which parameters are available for COSMO-SAC-Phi [78]; 'Binary' denotes the set of compounds used in the binary evaluation proposed by Jaubert et al. [94], which is discussed further below; and 'Complete' corresponds to the full set of molecules parameterized in the present study, also shown in Fig. 7. The full set of molecules appears twice: first, to enable comparison of the different datasets within openCOSMO-RS-Phi, and later, to facilitate comparison with other EoS.

To further illustrate model performance, ten of the 107 compounds in the 'Binary' dataset from Fig. 1 were selected to present the calculated vapor pressure and liquid molar volume. This selection covers pairings across all nine BACs analyzed later, while representing the largest possible number of binary systems (10.5 %) with this limited number of components. A visual comparison between experimental data and calculated results is shown in Fig. 2. The close agreement indicates the robustness and validity of the new method.

The corresponding numerical results are summarized in Table 3, where COSMO-SAC-Phi is denoted as 'CSP (2019)'. The table also includes a comparison with the improved version of COSMO-SAC-Phi containing one additional parameter, denoted as 'improved CSP (2021)' [81]. For the selected compounds, the new model outperforms the improved version without requiring a fifth adjustable parameter. It gives lower average and maximum errors in molar volume, lower values for both MAPE metrics, and greater consistency, as reflected by the smaller standard deviation. A complete table containing all pure components parameterized in this study is provided in the Supporting Information.

3.1.2. Derivative properties and critical points description

The preceding section presented a detailed assessment of the ability of openCOSMO-RS-Phi to reproduce liquid densities and saturated vapor pressures, i.e., the two properties used for the regression of the pure-component parameters. To provide a more comprehensive evaluation of the equation of state's performance, the accuracy for the molar enthalpy of vaporization (Δh_i^{LV}), liquid heat capacity at constant pressure evaluated at saturation ($c_{\mathrm{P,liq}}^{sat}$), critical temperature (T_c) and critical pressure (P_c) is examined in the following. The analysis is restricted to the 107 components included in the benchmark dataset of Jaubert et al. for mixtures.

Pure-component vaporization enthalpies. The enthalpy change on vaporization is identical to the difference between the molar residual enthalpy in the vapor ($h_{\mathrm{vap}}^{\mathrm{r}}$) and liquid ($h_{\mathrm{liq}}^{\mathrm{r}}$) phase of the pure component i in vapor–liquid equilibrium: $\Delta h_i^{\mathrm{LV}}(T, P_i^{sat}) = h_{\mathrm{vap},i}^{\mathrm{r}}(T, P_i^{sat}) - h_{\mathrm{liq},i}^{\mathrm{r}}(T, P_i^{sat})$. $\Delta h_i^{\mathrm{LV}}(T, P_i^{sat})$ denotes the molar enthalpy of vaporization of pure i at temperature T. One relevance of enthalpies of vaporization of pure components lies in the fact that, together with an accurate description of VLE's, they determine the accuracy of an EoS in predicting mixing enthalpies as eludicated in [102]. Similar to the procedure described in [104], a total of 50 enthalpies of vaporization

Table 3
Comparison between MAPE from openCOSMO-RS-Phi, improved COSMO-SAC-Phi, and COSMO-SAC-Phi, for a few selection of components.

Component	openCOSMO-RS-Phi				openCOSMO-RS-Phi		Improved CSP (2021)		CSP (2019)	
	b_i Å^3	δ^0_{m} $\frac{\text{kcal}}{\text{mol}}$	δ^t_m 10^2 K	$b_{h,i}$ Å^3	MAPE$_P$ %	MAPE$_v$ %	MAPE$_P$ %	MAPE$_v$ %	MAPE$_P$ %	MAPE$_v$ %
Methane	50.81	0.594	3.186	10.68	2.12	0.68	0.28	0.42	1.94	1.29
Ethane	73.49	0.803	3.549	12.68	1.78	0.67	0.47	0.72	1.48	2.30
Propane	99.34	0.845	3.572	14.66	1.24	0.63	0.33	0.49	1.01	2.29
Butane	126.31	0.860	3.657	15.66	0.67	0.54	0.18	0.43	0.51	1.16
Pentane	153.60	0.891	3.529	16.80	0.38	0.44	0.22	0.47	0.29	0.52
2-Methylpentane	181.83	0.887	3.444	17.75	0.22	0.45	2.24	1.66	2.10	4.33
Hexane	180.60	0.915	3.365	17.71	0.65	0.41	0.42	0.26	0.49	0.97
Octane	238.62	0.993	2.932	18.76	1.43	0.96	0.38	0.12	1.57	0.22
Decane	301.48	1.099	2.545	19.45	2.25	1.70	1.13	0.49	1.21	5.91
Cyclohexane	153.59	1.034	4.423	16.98	0.76	0.51	0.27	0.26	0.57	2.48
Ethene	65.57	0.667	3.960	11.95	1.75	0.66	0.14	0.40	1.40	1.02
1-Butene	117.13	0.865	3.534	15.36	1.37	0.65	0.41	0.50	0.51	0.55
1-Hexene	174.28	0.916	3.357	16.77	0.52	0.53	1.83	0.79	1.89	1.68
Cyclohexene	143.65	1.046	4.596	16.70	0.93	0.54			1.33	0.66
Benzene	126.16	1.083	4.782	15.88	1.00	0.55	0.38	0.36	0.59	1.81
Pyridine	117.23	1.203	4.998	16.09	0.51	0.57	0.71	2.43	0.71	1.50
Toluene	154.48	1.069	4.498	17.01	0.22	0.45	1.19	0.54	1.23	0.39
Acetone	98.69	1.048	3.982	15.87	0.22	0.44	0.70	0.73	0.80	0.85
2-Butanone	124.25	1.039	3.820	16.97	0.52	0.44			0.62	0.98
Dimethyl ether	82.90	0.914	3.787	13.29	0.99	0.62	0.25	0.18	0.64	2.60
Ethyl ether	137.91	0.926	3.276	16.19	0.56	0.48	0.30	0.55	0.39	0.79
Tetrahydrofuran	113.87	1.113	4.502	15.12	1.01	0.56	1.12	1.19	1.25	3.80
Methyl acetate	108.56	0.996	3.595	14.82	0.43	0.43	0.39	0.13	0.48	0.09
Ethyl acetate	137.42	1.017	3.146	15.98	0.38	0.45	0.41	0.86	0.48	1.71
Trichloromethane	111.60	1.062	4.491	14.18	0.97	0.54	1.11	0.87	1.31	0.74
Carbon tetrachloride	136.62	0.982	4.730	15.09	1.05	0.54	0.27	0.28	0.66	1.95
Methanol	56.09	1.351	3.325	8.71	0.57	0.38	0.55	0.24	0.82	3.45
Ethanol	85.74	1.541	2.258	9.14	1.73	1.29	0.81	0.17	0.74	4.63
1-Butanol	136.21	2.204	1.253	12.73	0.67	0.50	1.70	0.33	1.34	4.20
1-Pentanol	160.76	1.817	1.525	14.19	0.43	0.45	2.93	0.52	2.37	0.75
Methylamine	57.96	1.021	3.967	10.44	1.88	0.62	1.30	0.33	1.34	6.10
N,N-Diethylethanamine	199.01	0.904	3.464	17.55	0.70	0.57	1.86	0.07	1.14	2.18
Dimethylformamide	111.18	1.304	4.690	17.91	0.22	0.32	4.45	0.82	4.39	1.02
Acetic acid	80.19	0.768	6.236	11.84	0.55	0.23	2.57	6.40	2.02	6.72
Nitrogen	45.72	0.403	1.994	9.76	2.59	0.83	0.41	0.76	1.71	1.04
Carbon dioxide	46.92	0.886	3.331	9.21	0.18	0.15	0.11	0.16	0.19	0.18
Hydrogen sulfide	48.42	0.837	6.062	9.74	1.91	0.65	0.21	0.40	1.42	1.59
Ammonia	31.30	0.799	5.308	6.90	2.20	0.74	2.62	0.67	2.86	0.69
Dimethyl sulfoxide	107.43	1.451	4.859	16.84	0.48	0.61	1.21	1.62	0.94	1.88
Water	24.11	0.716	5.087	5.17	1.91	0.72	0.79	0.27	0.73	1.05
Average					1.00	0.59	0.96	0.73	1.19	1.95
Standard deviation					0.56	0.16	0.72	0.53	0.58	1.28

were calculated as reference data for each of the 107 components using the DIPPR equation 106 (correlation parameters for all components are included in the DIPPR database 801) by decomposing the temperature interval $[T^{\text{DIPPR}}_{\text{min}}, \min\{T^{\text{DIPPR}}_{\text{max}}; 0.98 \cdot T_{\text{c,exp}}\}]$ equidistantly. This results in a total of 5350 data points. Subsequently, Δh_i^{LV} was calculated for the same temperatures using the EoS, and a MAPE was computed for each component. The resulting MAPE distribution is shown in Fig. 3. Across all 107 components, a mean of 3.97 % and a median of 2.90 % are obtained. Removing acetic acid from the assessment, for which an unusually high MAPE of about 50 % is achieved, reduces the average to 3.53 %. Direct comparative values for the selected dataset are not available. For a larger number of components (1536), values of 1.92 % for tc-PR, 3.25 % for PC-SAFT, and 4.00 % for I-PC-SAFT are reported in [90].

Liquid isobaric molar heat capacities at saturation. The liquid isobaric molar heat capacity at saturation, $c^{sat}_{\text{P,liq}}$ is the sum of an ideal-gas and residual contribution: $c^{\text{sat}}_{\text{P,liq}}(T, P^{sat}) = c^{\text{IG}}_{\text{P}}(T) + c^{\text{r}}_{\text{P,liq}}(T, P^{sat})$. Since an EoS only predicts $c^{\text{r}}_{\text{P,liq}}$, an accurate temperature-dependent model for the ideal gas molar heat capacity has to be coupled with the EoS. Here, a DIPPR correlation for ideal-gas heat capacity was employed. As for the evaluation of Δh_i^{LV}, for each of the 107 components, $c^{sat}_{\text{P,liq}}$ was calculated for 50 temperature points on the temperature domain $\max\{T^{\text{DIPPR}}_{\text{min},c_{\text{P,IG}}}; T^{\text{DIPPR}}_{\text{min},c_{\text{P,liq}}}\}, \min\{T^{\text{DIPPR}}_{\text{max},c_{\text{P,IG}}}; T^{\text{DIPPR}}_{\text{max},c_{\text{P,liq}}}; 0.98{\cdot}T_c\}$ and compared with the corresponding values obtained from the DIPPR correlations. The MAPE was again used as the performance metric. The resulting MAPE values are shown in Fig. 4a. The MAPE ranges from 0.60 % for butane to 31.48 % for carbon tetrafluoride, which is comparable to the spread reported for classical and ML-based tc-PR variants [105], although the present evaluation is based on a smaller dataset. The mean and median MAPE values are 10.3 % and 8.7 %, respectively. Fig. 4b compares the predicted and reference saturated liquid molar heat capacities for seven selected real fluids. Finally, it should be pointed out that it is well-known that the obtained MAPE value depends on the chosen fitting strategy, particularly on whether energetic properties are included in the parameter regression, as [104] shows using PC-SAFT as an example.

Critical properties. To assess the performance of openCOSMO-RS-Phi in critical-point calculations, Fig. 5 shows parity plots comparing predicted critical temperatures (T_c) and critical pressures (P_c) with experimental DIPPR data for all pure components included in the database of Jaubert et al.

The resulting MAPE values are 3.32 % for T_c and 22.98 % for P_c, indicating substantially larger deviations for critical pressure. The largest errors in both properties are found for the long-chain alkanes

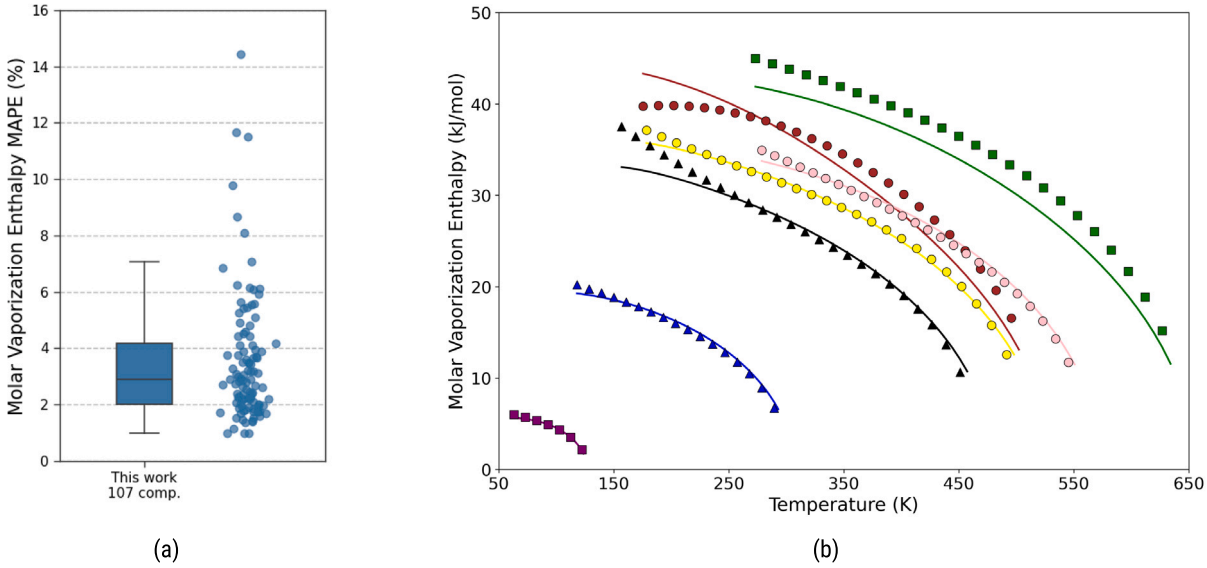

Fig. 3. (a) Resulting MAPE for the molar enthalpy change of vaporization, evaluated across the 107 components of the benchmarking database. Acetic acid with a MAPE > 50 % is excluded from the scatter plot. (b) Temperature dependence of the vaporization enthalpy for seven selected components: • Acetone, • Methanol, ▲ Diethyl ether, • Benzene, ▲ Trifluoromethane, ■ Water, ■ Nitrogen. Points are DIPPR data, and lines are values given using openCOSMO-RS-Phi. For the sake of clarity, not all 50 DIPPR points used are shown.

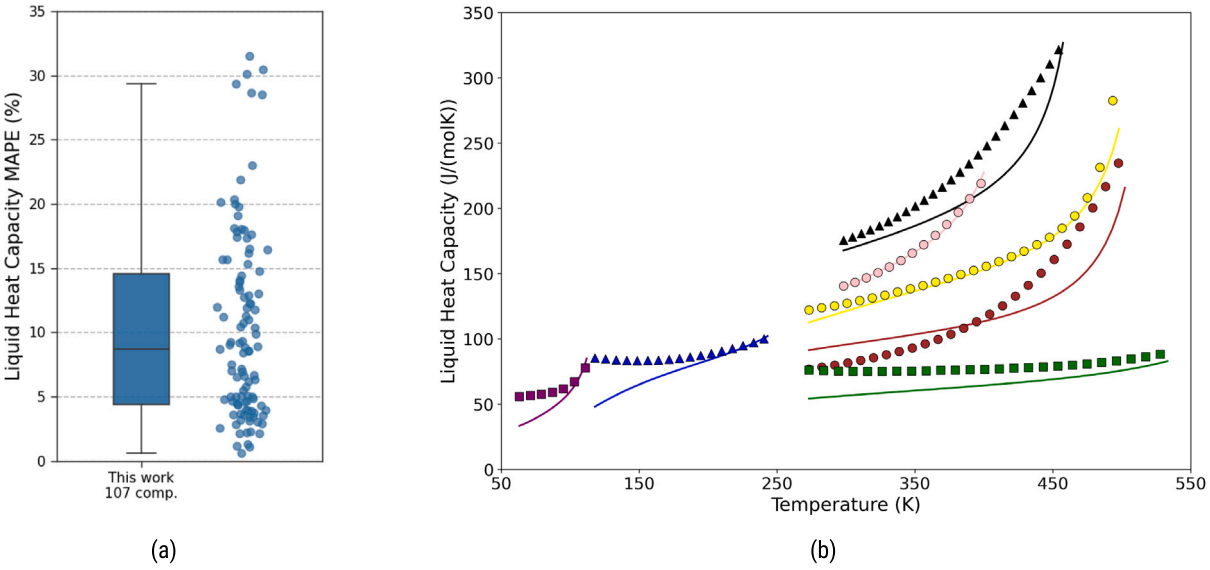

Fig. 4. (a) Resulting MAPE for isobaric molar heat capacities of liquids at saturation, $c_{P,liq}^{sat}$, evaluated across 107 components. The six compounds exhibiting the largest prediction errors, all with MAPE values of approximately 30 %, are (in descending order of MAPE): carbon tetrafluoride, xenon, chlorodifluoromethane, dichloromethane, ethane and cyclopentane. (b) Temperature dependence of the molar liquid heat capacity for seven selected components: • Acetone, • Methanol, ▲ Diethyl ether, • n-Butane, ▲ Trifluoromethane, ■ Water, ■ Nitrogen. Points are DIPPR data, and lines are values given using openCOSMO-RS-Phi. For the sake of clarity, not all 50 DIPPR points used are shown.

n-hexatriacontane, *n*-dotriacontane, and *n*-tetracosane, which appear as clear outliers in Fig. 5. Overall, T_c is overestimated for most compounds, with ammonia as the only exception, whereas P_c is systematically overestimated for all components.

A similar trend has been reported for other EoS [106]. For example, PC-SAFT and CPA have been reparameterized to improve agreement with T_c and P_c [107,108], and it has also been shown that fitting PR parameters to vapor-pressure and density data, rather than deriving them from critical properties, leads to poorer reproduction of critical points [109].

3.1.3. Comparison with other EoS

The performance of openCOSMO-RS-Phi is further evaluated using the same analysis previously applied to other EoS [90,91]. As shown in Fig. 6, the assessment considers both the complete dataset and its subdivision into non-self-associating (NSA) and self-associating

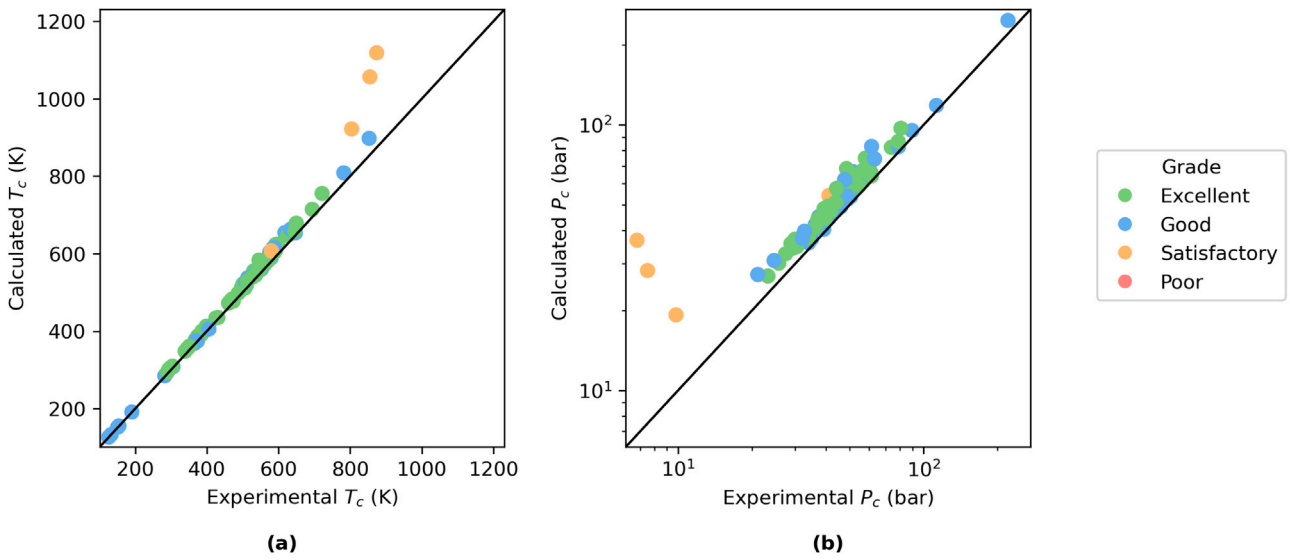

Fig. 5. A parity plot of (a) the predicted critical temperature (K) from openCOSMO-RS-Phi compared to experimental critical temperature (K) obtained from DIPPR on a log–log scale and (b) the predicted critical pressure (bar) from openCOSMO-RS-Phi compared to experimental pressure temperature (bar) obtained from DIPPR. The three clearly visible outliers refer to hexatriacontane, n-dotriacontane, and n-tetracosane.

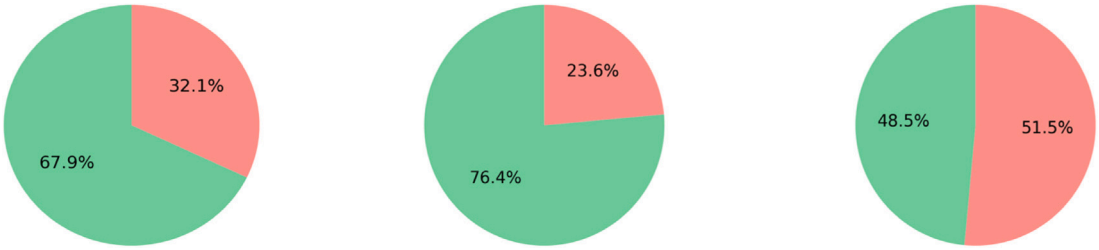

Fig. 6. Ratio of • well- and • badly-modeled molecules for the entire dataset, non-self-associating (NSA), and self-associating (SA) fluids.

(SA) components. Consistent with the studies of Piña-Martinez et al. and Ramírez-Vélez et al., openCOSMO-RS-Phi shows lower internal performance for SA fluids than for NSA fluids.

However, the ratio of well-modeled to badly modeled compounds should not be interpreted directly as a measure of relative model performance. This classification is inherently model dependent and reflects only the internal consistency of a given EoS. Its value for comparisons between different models is therefore limited and should be interpreted with caution. A more detailed distribution of the MAPE for each property is provided in Appendix A3 of the Supporting Information.

The MAPE values for vapor pressure and molar volume across all components are presented in Fig. 7, allowing a direct comparison of the overall performance of openCOSMO-RS-Phi with that of other EoS. For vapor pressure, all models show comparable behavior and generally satisfactory accuracy. For molar volume, in contrast, clearer differences between the models are observed.

For molar volume, both PC-SAFT and openCOSMO-RS-Phi perform reasonably well; however, PC-SAFT shows higher accuracy in terms of both mean error and error distribution. This difference may be related to the treatment of repulsive interactions in openCOSMO-RS-Phi, which relies on the classical van der Waals formulation. Such an approach may limit predictive accuracy, particularly for larger molecules. A direct comparison between the two models is provided in Fig. 8.

To further illustrate the capabilities of the EoS across different molecular groups, the proposed accuracy-based classification is applied to the chemical families defined in the DIPPR database. The performance classification is based jointly on the MAPE values obtained for vapor pressure and liquid molar volume. A compound is assigned to a given class only if the errors for both properties satisfy the corresponding threshold, such that the classification is determined by the less accurately represented property. Compounds are classified as *excellent* when both MAPEs are less than or equal to 1 %, *good* when both are less than or equal to 5 %, *satisfactory* when both are less than or equal to 10 %, and *poor* when either MAPE exceeds 10 %. This classification has previously been used for PC-SAFT [91] and tc-PR [90], and is extended here to openCOSMO-RS-Phi, as summarized in Table 4. In this table, the abbreviations comp., exc., and sat. denote compounds, excellent, and satisfactory, respectively.

As shown in Table 4, the trends are consistent with those in Fig. 7. The tc-PR model predominantly yields errors in the range of 1 to 5 % and is therefore mainly classified as good, whereas PC-SAFT achieves errors largely below 5 %, spanning both the good and excellent categories. openCOSMO-RS-Phi shows a substantial proportion of compounds classified as excellent and good, together with a smaller fraction categorized as satisfactory. This behavior can again be linked to the reduced accuracy of openCOSMO-RS-Phi for larger molecules, as illustrated in Fig. 8.

3.1.4. Adjusted parameter evaluation

An examination of the correlations between the adjusted parameters and molecular size reveals trends similar to those reported for COSMO-SAC-Phi. In particular, a linear relationship is observed between the adjusted parameter b_i and the COSMO volume, as shown in Fig. 9(a). Fig. 9(b) presents the corresponding relationship between the hole molar cavity, representing free volume, and the COSMO volume. Distinct regions can be identified. For the parameter b_i, the molecular volume

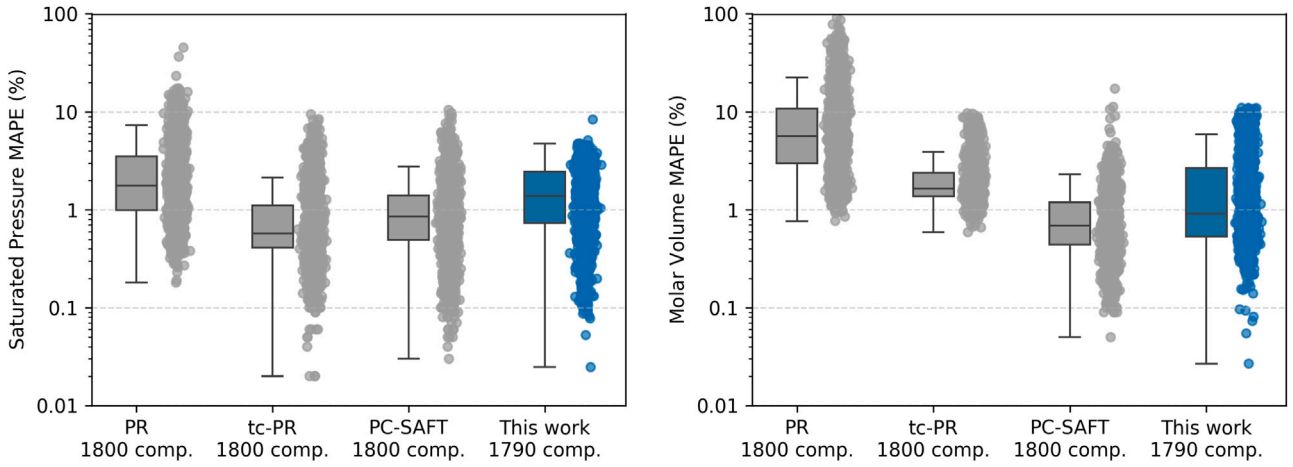

Fig. 7. Comparison of the resulting MAPE of vapor pressure and molar volume from all components studied across different EoS.

Table 4
Accuracy classification based on vapor pressure and molar volume predictions for compounds grouped by chemical family using tc-PR, PC-SAFT, and openCOSMO-RS-Phi. The classification system is explained in the text.

Chemical family	Number of	tc-PR				PC-SAFT				openCOSMO-RS-Phi			
	comp.	Exc.	Good	Sat.	Poor	Exc.	Good	Sat.	Poor	Exc.	Good	Sat.	Poor
Alcohols	130	2	119	9	0	13	110	7	0	39	59	29	3
Alkanes	163	3	158	2	0	131	32	0	0	89	64	10	0
Alkenes	122	0	121	1	0	94	28	0	0	72	45	4	1
Alkynes	18	1	15	2	0	12	6	0	0	7	11	0	0
Amines/amides	103	2	95	6	0	43	59	1	0	28	59	16	0
Aromatics	150	1	147	2	0	75	75	0	0	35	105	10	0
Esters/ethers	238	0	228	10	0	112	124	2	0	60	109	60	9
Halogen compounds	232	9	214	9	0	112	118	2	0	130	95	7	0
Inorganic compounds	72	0	58	14	0	24	41	5	2	19	44	1	0
Ketones/aldehydes	80	2	77	1	0	35	45	0	0	31	47	2	0
Nitriles	31	1	23	7	0	1	27	2	1	9	18	4	0
Nitrogen compounds	98	5	86	7	0	22	72	4	0	28	54	16	0
Organic acids	82	4	76	2	0	11	71	0	0	4	32	44	2
Organic salts	13	0	8	5	0	5	7	1	0	3	8	2	0
Other compounds	32	0	30	2	0	16	15	1	0	9	23	0	0
Peroxides	11	0	10	1	0	3	8	0	0	2	4	5	0
Polyfunctional compounds	70	1	63	6	0	24	43	3	0	7	42	21	0
Silicon compounds	61	2	54	5	0	22	39	0	0	19	26	13	1
Sulfur compounds	94	2	90	2	0	66	28	0	0	40	53	1	0
Total	1800	35	1672	93	0	821	948	28	3	631	898	245	16

correlation seems to hold true for most compounds. However, for the parameter b_h this correlation is much weaker.

The model shows limitations in accurately describing large molecules and the effects associated with them. In particular, the combination of increasing molecular size and high mean absolute percentage error (MAPE) in molar-volume predictions leads to lower values of b_h. This behavior can be attributed to the selected hard-sphere term, which relies on a simplified representation of repulsive interactions. Such an approach becomes less suitable for larger and more structurally complex molecules, for which size, shape, and packing effects are not adequately captured.

A more detailed analysis of the variables contributing to this behavior is provided in Figure SI14 in the Supporting Information, complementing the trends already discussed in Figs. 8 and 9(b).

No clear global correlation is observed for δ_m^0 and δ_m^t. These parameters are plotted against COSMO volume in Figure SI15 in the Supporting Information, where the influence of local minima, previously described in the three-step minimization approach, becomes evident. This behavior is consistent with the role of δ_m^0 and δ_m^t as effective fitting parameters, since they absorb contributions from molecular effects that openCOSMO-RS cannot describe explicitly. In this sense, all phenomena not properly captured by the model become embedded in these parameters during the fitting procedure.

Nevertheless, clear trends emerge within specific chemical subgroups for δ_m^0 and δ_m^t, as shown in Figure SI16 and Figure SI17 in the Supporting Information for *n*-alcohols and *n*-alkanes, respectively, as a function of COSMO volume. For both groups, moderately accurate functional relationships can be identified, providing a useful description of the behavior of the systems under investigation.

3.2. Binary mixtures

As outlined in the Methodology section, the performance of openCOSMO-RS-Phi for binary-mixture thermodynamic data is assessed using the grading procedure proposed by Jaubert et al. [94], including the modifications introduced by Piña-Martinez et al. [95]. The corresponding results are presented in this section, beginning with an overview and a comparison with other models. Since the performance of the EoS depends on both the property and the BAC of the system considered, a more detailed analysis is carried out for each BAC and property. Representative example plots are included to illustrate model

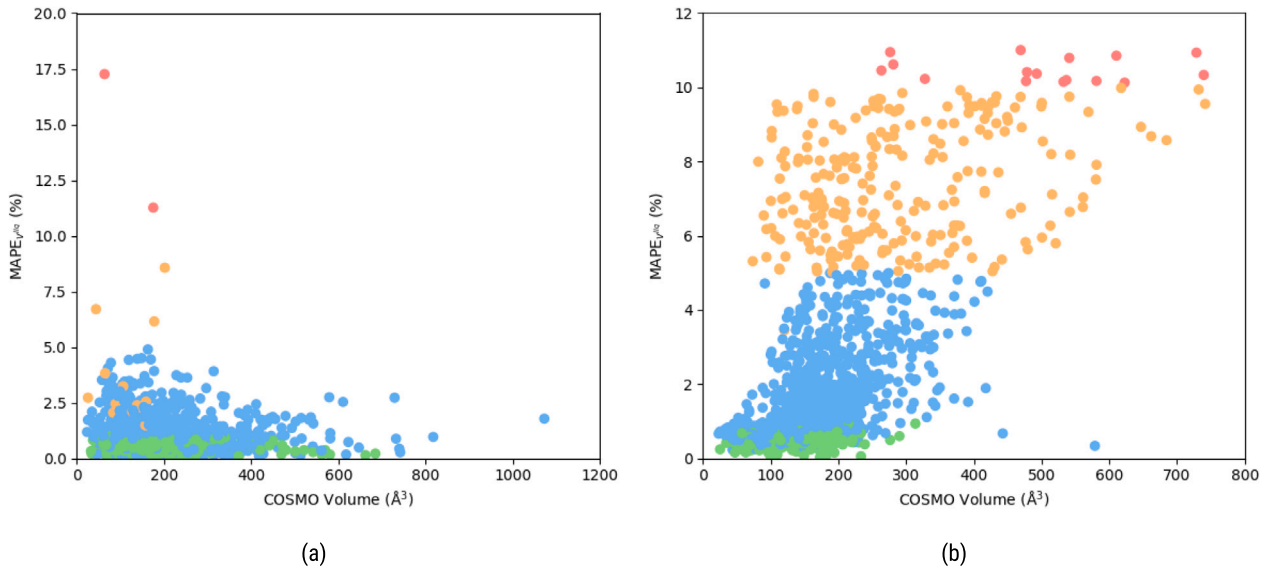

Fig. 8. Correlation between COSMO volume and molar-volume MAPE for (a) PC-SAFT and (b) openCOSMO-RS-Phi, with points grouped according to an accuracy-based classification. • excellent, • good, • satisfactory, • poor.

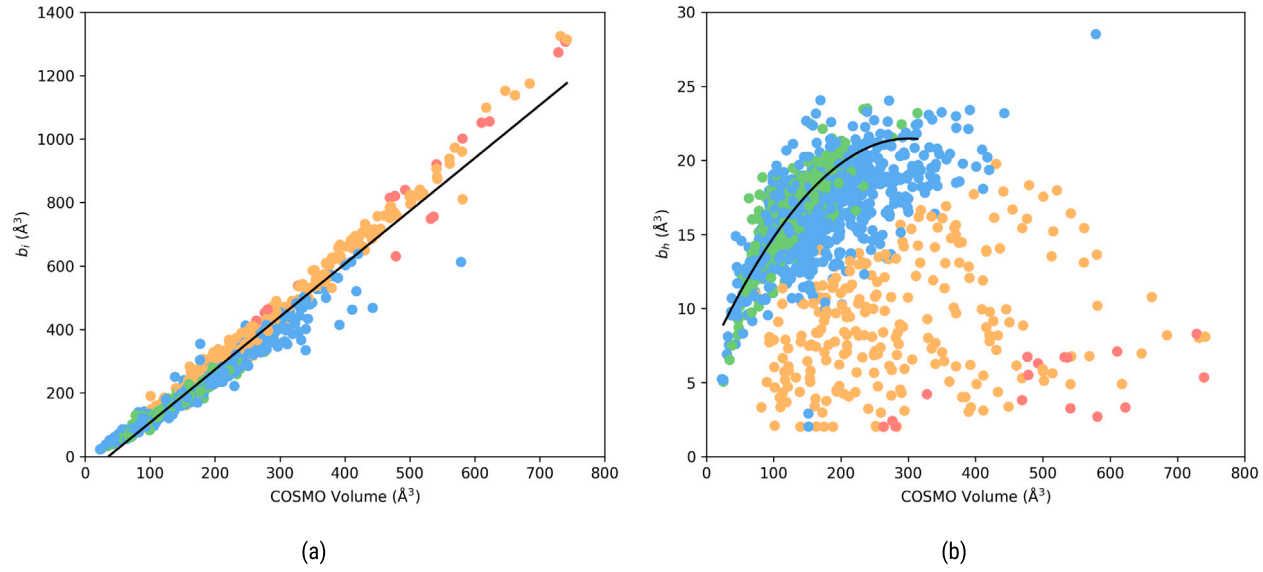

Fig. 9. Molar cavity (a) and hole molar cavity (b) as a function of COSMO volumes. • excellent, • good, • satisfactory, • poor.

behavior. For the sake of comparability with previous benchmark studies, only systems that were also illustrated in those studies are plotted.

3.2.1. Overall benchmark

Within the applied benchmark procedure, the model performance is quantified by the MAPE and the SR, which accounts for OM points. Table 5 reports the MAPEs for ten properties — liquid-phase composition (x), vapor-phase (or second liquid-phase) composition (y), three-phase pressure (P_{LLV}), three-phase composition (z_{LLV}), critical pressure (P_c), critical composition (x_c), azeotropic pressure (P_{azeo}), azeotropic composition (x_{azeo}), enthalpy of mixing (h^M), and heat capacity of mixing (c_P^M) — for each BAC, together with the corresponding SR values.

These MAPE and SR values are subsequently converted into marks (score over 20) for each BAC and property. Then, Mark20 is calculated by category of association (N-A, S-A, C-A, S-A + C-A). The overall score is obtained as the average of the four category scores. For openCOSMO-RS-Phi, a final score of 7.3/20 is obtained, corresponding to fair performance. The resulting marks are summarized in Table 6.

By penalizing low SR values, the modified grading scheme provides a more realistic measure of predictive capability than the original procedure. This is especially relevant for openCOSMO-RS-Phi, since several SR values are low (see Table 5).

For example, pronounced difficulties are observed for VLLE and azeotropic data of BAC$_8$. When SR is taken into account, the marks decrease from 17.1 (x_{LLV}) and 19.7 (P_{LLV}) to 1.4 and 1.6, respectively, and from 3.3 (x_{azeo}) and 13.7 (P_{azeo}) to 1.1 and 4.6.

For a more comprehensive assessment of openCOSMO-RS-Phi, Fig. 10 compares its global mark with those of other thermodynamic models benchmarked against the database of Jaubert et al. using the same methodology and evaluation criteria. These models include

Table 5
Overview of the MAPE between Experimental Data and Model Calculations along with the Corresponding SRs (openCOSMO-RS-Phi without BIPs).

BAC	MAPE and corresponding SR on:													
	x	SR_x	y	SR_y	P_{LLV}	z_{LLV}	SR_{LLV}	P_c	x_c	SR_c	P_{azeo}	x_{azeo}	SR_{azeo}	h^M	c_P^M
1 (NA-NA)	23.9	0.93	13.8	0.93	3.3	17.5	0.5	22.8	38.3	0.81	4.1	4.6	1.0	42.9	88.9
2 (HA-NA)	25.8	0.81	18.1	0.81				28.0	49.0	0.65	8.9	27.6	0.68	46.8	199.6
3 (HD-NA)	40.0	0.68	23.1	0.68				14.5	49.4	0.57	21.0	35.8	0.41	52.2	117.9
4	19.4	0.90	16.2	0.90				21.9	53.0	0.87	5.3	26.0	0.29	31.0	196.9
5 (SA-NA)	44.3	0.64	14.7	0.64	3.7	23.6	1.0	19.6	85.4	0.47	4.4	12.2	1.0	45.3	61.9
6 (HD-HA)	36.3	0.90	30.2	0.90				11.8	60.1	0.83	10.9	18.8	1.0	40.9	86.6
7 (SA-HD)	39.3	0.70	21.7	0.70							3.9	22.7	0.6	65.6	49.3
8 (SA-HA)	47.0	0.76	22.4	0.76	0.5	5.9	0.08	20.5	43.1	0.76	12.5	33.4	0.33	61.7	88.9
9 (SA-SA)	33.0	0.91	21.8	0.91	7.1	6.0	1.0	29.8	55.6	0.77	6.9	20.3	1.0	57.3	75.4

Table 6
Rating of the nine BACs To finally grade the model (openCOSMO-RS-Phi with no BIPs).

BAC	Mark on:										BAC mark	Category mark	Final mark
	x	y	P_{LLV}	z_{LLV}	P_c	x_c	P_{azeo}	x_{azeo}	h^M	c_P^M			
1 (NA-NA)	7.5	12.2	9.2	5.6	2.3	0.7	18.0	17.7	9.3	11.1	9.4		
2 (HA-NA)	5.7	8.9			0.0	0.0	10.6	4.2	8.3	0.0	4.7	5.8	
3 (HD-NA)	0.0	5.8			5.2	0.0	3.9	0.9	7.0	8.2	3.9		
4 (HD/HA-HD/HA)	9.3	10.7			3.1	0.0	5.0	2.0	12.2	0.3	5.3		
5 (SA-NA)	0.0	8.1	18.2	8.2	2.5	0.0	17.8	13.9	8.7	13.8	9.1	9.1	**7.3**/20
6 (HD-HA)	1.6	4.4			9.2	0.0	14.5	10.6	9.8	11.3	7.7	7.7	
7 (SA-HD)	0.3	6.4					10.8	5.2	3.6	15.1	6.9		
8 (SA-HA)	0.0	6.7	1.6	1.4	3.5	0.0	4.6	1.1	4.6	11.1	3.5	6.4	
9 (SA-SA)	3.2	8.3	16.5	17.0	0.0	0.0	16.6	9.8	5.7	12.5	8.9		
Overall	2.1	7.3	12.1	7.7	4.0	0.0	13.1	9.0	8.1	10.7			

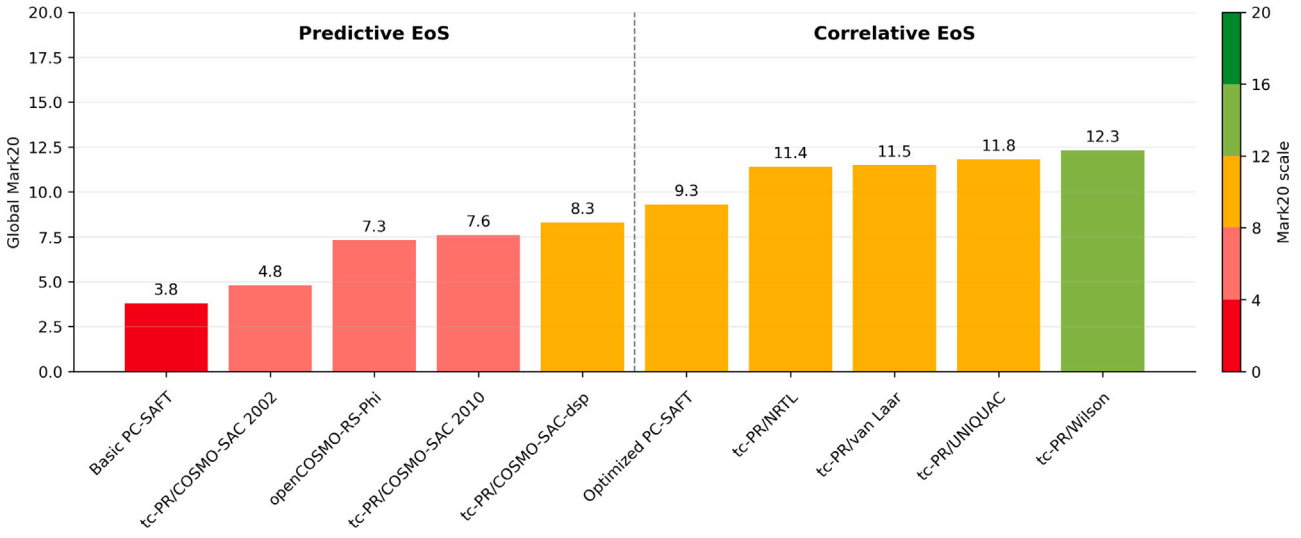

Fig. 10. Comparison of the global marks of different EoSs following the methodology proposed in [94,95].

predictive EoS without binary interaction parameters (BIPs), such as basic PC-SAFT [97] and tc-PR combined with various COSMO-SAC versions [96], as well as correlative EoS with BIPs regressed to experimental data, such as optimized PC-SAFT [98] and tc-PR coupled with activity-coefficient models including Wilson, NRTL, UNIQUAC, and van Laar [95]. It should be noted that, for tc-PR combined with COSMO-SAC-2002, COSMO-SAC-2010, and COSMO-SAC-dsp, the calculations were performed for only 160 out of the 200 systems. Consequently, comparisons with these models are not exact. In addition to the global scores shown in Fig. 10, detailed comparisons of all models in terms of individual properties are provided in the Supporting Information.

Among the predictive EoS, openCOSMO-RS-Phi performs significantly better than basic (pure-component) PC-SAFT and tc-PR combined with COSMO-SAC-2002. However, it performs worse than tc-PR combined with COSMO-SAC-2010 and tc-PR-COSMO-SAC-dsp, primarily because of a less accurate description of two-phase compositions (x, y) and critical compositions (x_c). While openCOSMO-RS-Phi employs the COSMO-RS framework to compute the attractive pressure through a pseudo-mixture approach, tc-PR-COSMO-SAC-dsp incorporates the COSMO-based activity-coefficient model only into the equation system used to determine the mixture attraction parameter a. Other differences that may affect performance include the global parameters, the

determination of pure-component parameters (regression versus estimation from critical data), the absence of volume translation in openCOSMO-RS-Phi, differences in the mixing rules for the co-volume parameter b_i, and variations in the temperature dependence of the interaction-energy contributions.

As shown in Fig. 10, predictive models generally perform worse than correlative EoS, which can be attributed to the absence of binary interaction parameters. However, this limitation is offset by their broader applicability, including the ability to predict previously unseen mixtures. For systems with sufficient experimental data, the introduction of binary parameters, for example in the dispersion term as proposed by Soares et al. [78], is expected to significantly improve performance. For example, incorporating k_{ij} values and induced association parameters increased the PC-SAFT score from 3.8 to 9.3 [98]. Further potential for improvement arises from the aspects discussed above. The aim of this work was to benchmark a fully predictive base version of openCOSMO-RS-Phi and thereby provide a reference for future developments of the EoS.

Benchmark across BACs. At the BAC level, the highest score is obtained for the BAC_1 group (9.4/20), which comprises mixtures of two non-associating compounds. However, because of the lower scores for BAC_2–BAC_4, the aggregated non-associating (N-A) category (BAC_1–BAC_4) represents the weakest-performing category overall. For non-associating systems with at least one non-polar component, the relative importance of the dispersion contribution is higher than that of the misfit and hydrogen-bonding terms. The comparatively poor performance may therefore be attributed to the fact that the global parameters of openCOSMO-RS-Phi were regressed without explicitly accounting for dispersion contributions, which are included only in the EoS formulation. In addition, Zini et al. reported that the dispersion term used in this implementation lacks a clear theoretical foundation and therefore proposed an alternative term [81].

The second-highest BAC-level score, and the highest category score after averaging, is obtained for the BAC_5 group, with $mark_{SA}$ = 9.1. This favorable result is mainly due to the accurate representation of three-phase line data, azeotropic behavior, and energetic properties. Notably, this score is the second-highest reported for BAC_5 among all predictive and correlative EoS evaluated to date. The BAC_5 group consists of mixtures of non-associating and self-associating components, such as alkane + alcohol systems. Interestingly, basic PC-SAFT also shows its best performance in this category, which has been attributed to its explicit treatment of hydrogen-bonding interactions. A similar explanation likely applies to openCOSMO-RS-Phi: in the presence of self-associating components, the relative importance of dispersion interactions decreases, favoring models with an accurate description of association effects.

The second-best category score is obtained for the cross-associating (C-A) group, comprising BAC_6 systems. Here, openCOSMO-RS-Phi performs significantly better than PC-SAFT and tc-PR/COSMO-SAC-2002, but worse than tc-PR/COSMO-SAC-2010 and tc-PR/COSMO-SAC-dsp. These systems are typically characterized by negative deviations from ideality. Paes et al. showed that inclusion of a dispersion term reduces systematic errors for BAC_6 systems [96]. A more accurate representation of hydrogen bonding further improves model performance. In the present implementation of openCOSMO-RS-Phi, hydrogen bonding is described using a simplified approach with a single global c_{hb} parameter and no temperature dependence. A more refined treatment of hydrogen bonding may therefore improve the results further.

The results for the fourth category, comprising mixtures with both self- and cross-association (BAC_7–BAC_9), are mixed. While openCOSMO-RS-Phi shows the best performance among predictive EoS for BAC_9 systems, it performs comparatively poorly for BAC_7 and BAC_8. The lowest score across all BACs is obtained for BAC_8, with a value of 3.5/20. For most EoS benchmarked to date, however, the lowest performance has typically been observed for BAC_6.

Finally, it is noteworthy, and expected, that similar qualitative trends, such as the comparatively good description of complex systems involving full or partial self-association, were also reported by Nikolaidis et al. in the benchmark of PC-SAFT [97]. This can be attributed to the fact that both approaches explicitly account for hydrogen-bonding interactions, in contrast to cubic EoS such as Peng–Robinson.

3.2.2. Phase equilibria

Two-phase data. For the prediction of liquid-phase compositions ($\mathbf{x}$) at given P and T, the lowest MAPE is obtained for BAC_4 (19.4 %), whereas the highest MAPE is observed for BAC_8 (47.0 %), indicating the poorest performance in this class. For vapor-phase compositions ($\mathbf{y}$), the lowest MAPE is achieved for BAC_1 (13.8 %), while the highest value (30.2 %) is found for BAC_6.

The SR for two-phase equilibria ranges from 64 % to 93 %, with an overall average of 0.83. These values are comparable to those reported for other predictive EoS (e.g., PC-SAFT: 74–80 %, PR: 78 %, SAFT-VR Mie: 85 % [97,110]), whereas correlative models typically reach around 95 % [95]. Reduced SR values, particularly for BAC_5, are a common limitation of predictive approaches and reflect difficulties in reproducing the correct qualitative phase behavior.

For openCOSMO-RS-Phi, these deficiencies are mainly linked to an underestimation of azeotropic pressures, especially for positive azeotropes. Despite this, the phase diagrams are generally reproduced qualitatively well. The P–x–y diagrams, including the VLE region, predicted by openCOSMO-RS-Phi for 15 different binary systems are presented in Fig. 11.

Azeotropic data. openCOSMO-RS-Phi achieves a score of 13.1 for azeotropic pressure (P_{azeo}), indicating overall good performance. The corresponding MAPEs range from 3.9 % (BAC_7) to 21.0 % (BAC_3). The deviations arise almost exclusively from a systematic underestimation of azeotropic pressures; overestimations are rare and occur only for a few BAC_9 systems. SR values span a wide range from 0.29 to 1.0.

Excellent performance is observed for mixtures belonging to BAC_1 (18.0/20), BAC_5 (17.8/20), and BAC_9 (16.6/20), all with SR = 1. For BAC_9, openCOSMO-RS-Phi yields the best result among all EoS evaluated to date. In contrast, the lower scores for BAC_3, BAC_4, and BAC_8 are primarily caused by reduced SR values.

However, the statistical significance of these results is limited by the relatively small azeotropic dataset (225 data points across nine BACs). In particular, the mark for BAC_4 is based on only seven data points, five of which correspond to the dimethyl ether + sulfur dioxide system at different temperatures. For this system, openCOSMO-RS-Phi predicts nearly ideal behavior instead of azeotropy, resulting in a low SR (29 %) and a substantial reduction of the overall score despite good performance for the remaining systems.

The scores for azeotropic compositions (x_{azeo}) are consistently lower than those for P_{azeo}, in line with observations for other EoS. The best performance is again obtained for BAC_1, BAC_5, BAC_6, and BAC_9, with openCOSMO-RS-Phi achieving the highest reported score for BAC_9 (9.8). Larger deviations persist for BAC_3, BAC_4, and BAC_8. Eight systems with azeotropic behavior are presented in Fig. 11(c), (d), (e), (f), (g), (j), (k), and (n).

VLLE. Three-phase line data are available for BAC_1, BAC_5, BAC_8, and BAC_9 systems. For both three-phase pressure (P_{LLV}) and composition (z_{LLV}), openCOSMO-RS-Phi performs significantly better than most predictive EoS and reaches a level comparable to that of correlative models.

A notable exception is the very low SR value for BAC_8 (8 %). This is mainly due to the composition of the dataset: 11 out of 12 experimental three-phase lines correspond to the dimethyl ether + water system, for which openCOSMO-RS-Phi fails to predict heteroazeotropic behavior. Although the remaining system, CO_2 + water at 298.15 K, is described

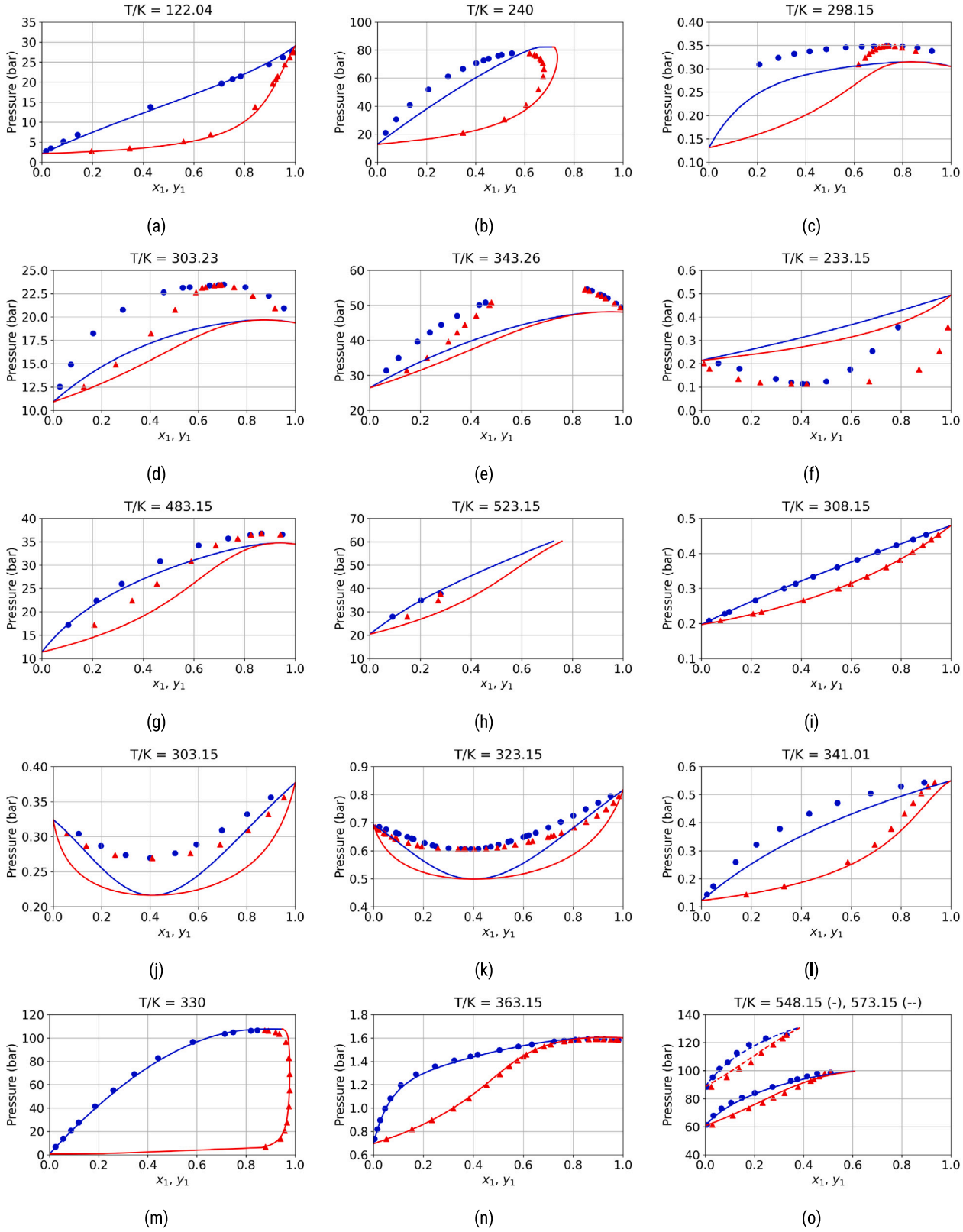

Fig. 11. P–x–y diagrams for binary mixtures exhibiting vapor–liquid equilibrium at the specified temperature, calculated with openCOSMO-RS-Phi: •: x_1^{exp}, ▲: y_1^{exp}, ——: x_1^{calc}, ——: y_1^{calc}; (a) nitrogen (1) - methane (BAC_1); (b) methane (1) - carbon dioxide (BAC_2); (c) acetone (1) - cyclohexane (BAC_2); (d) and (e) difluoromethane (1) - propane (BAC_3); (f) dimethyl ether (1) - sulfur dioxide (BAC_4); (g) and (h) ethanol (1) - *n*-heptane (BAC_5); (i) diethylamine (1) - benzene (BAC_5); (j) and (k) acetone (1) - chloroform (BAC_6); (l) trichloroethylene (1) - 2-methoxyethanol (BAC_7); (m) carbon dioxide (1) - methanol (BAC_8); and (n) and (o) ethanol (1) - water (BAC_9).

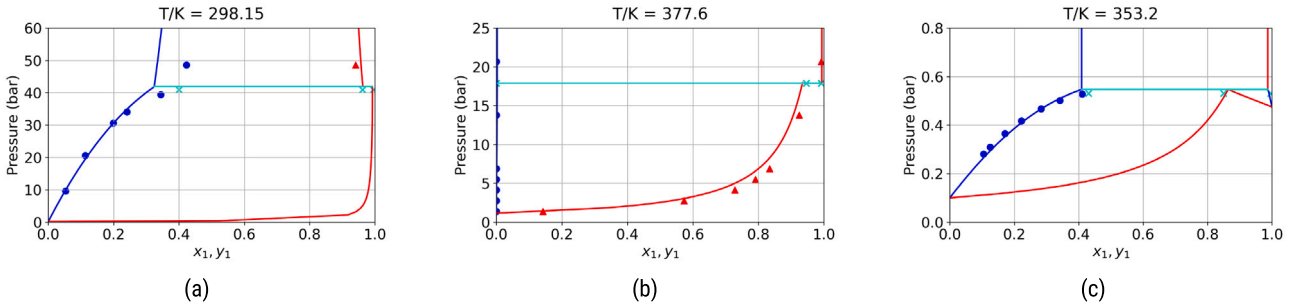

Fig. 12. P–x–y diagrams for three binary mixtures that exhibit a VLLE for the specified temperature calculated with openCOSMO-RS-Phi: •: x_1^{exp} / $x_1^{\alpha,exp}$, ▲: y_1^{exp} / $x_1^{\beta,exp}$, ×: z_{LLV}^{exp}, —: x_1^{calc} / / $x_1^{\alpha,calc}$, —: y_1^{calc} / / $x_1^{\beta,calc}$, —: calculated three-phase line; (a) ethane (1) - methanol (BAC_5), (b) *n*-butane (1) - water (BAC_5), (c) water (1) - 1-pentanol (BAC_9).

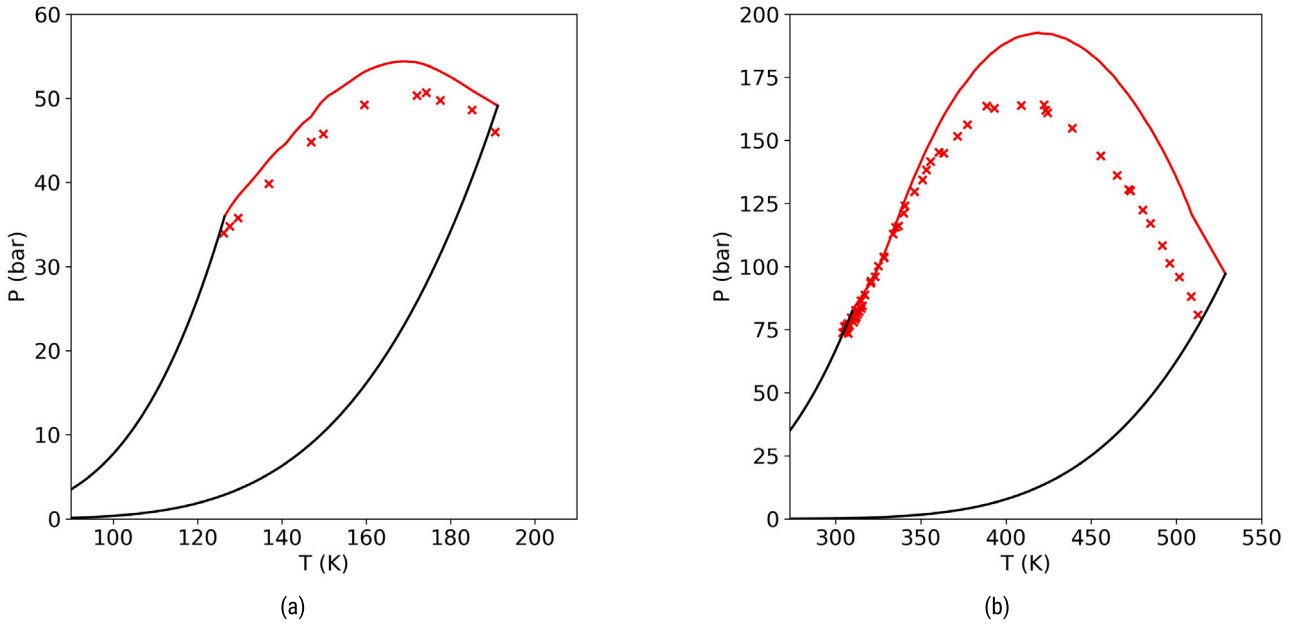

Fig. 13. Global phase equilibrium diagram (GPED) for two binary systems calculated with openCOSMO-RS-Phi: (a) N_2 (1) - methane (BAC_1), (b) carbon dioxide (1) - methanol (BAC_8). It includes the vapor-pressure curves of the pure components (black lines), experimental critical points (×) and the calculated critical line (red line). The critical line is evaluated on a temperature grid with $\Delta T = 2$ K and mixture critical points are determined using a tracing procedure as outlined in Section 2.5.

very accurately, the overall result is dominated by dimethyl ether + water.

In Fig. 12, P–x–y diagrams including both the VLE and LLE regions predicted by openCOSMO-RS-Phi for three binary systems are presented.

Critical data. A clear weakness of openCOSMO-RS-Phi is the description of mixture critical points. For the critical composition (x_c), the overall score (Mark20) is 0.0, resulting from MAPEs above 40 % for all BACs except BAC_1.

A similar trend is observed for the critical pressure (P_c), although openCOSMO-RS-Phi performs somewhat better than the other predictive EoS, achieving a score of 4.0. The large deviations are mainly caused by an inaccurate representation of the bubble and dew curves near the critical region, typically leading to an overestimation of the critical pressure. This behavior is illustrated in Fig. 13, which shows the critical line for the non-associating binary system N_2 + CH_4.

In addition to the MAPE, the SR values provide further insight. For BAC_5 systems, only 47 % of the critical points are reproduced, whereas BAC_4 shows the highest SR (87 %). Although openCOSMO-RS-Phi attains slightly higher average SR values than PC-SAFT or COSMO-based tc-PR EoS, these values still indicate significant deficiencies. As seen in Fig. 5, the model systematically overestimates critical temperatures of pure components, such that no critical point is predicted for mixtures at temperatures slightly above $\min(T_{c,1}, T_{c,2})$.

3.2.3. Derivative properties

Derivative properties reflect the temperature dependence of a thermodynamic model. For the enthalpy of mixing (h^M) and heat capacity of mixing (c_P^M), openCOSMO-RS-Phi shows notably strong performance. In particular, it outperforms all other predictive EoS for h^M and achieves the best overall score for c_P^M (10.7), even exceeding all predictive and correlative models, although some room for improvement remains. In addition, it should be noted that the marks reported in Table 6 are based on the asymmetric MAPE according to Eq. (39) which might differ from other benchmark studies. For a more detailed explanation of this topic, the reader is referred to the SI. Even with the alternative use of a symmetrical MAPE, openCOSMO-RS-Phi still shows better results than other predictive EoS models.

This result is noteworthy, since COSMO-based models are often considered less reliable for temperature-dependent and energetic properties, and such data were not used in the parameterization of

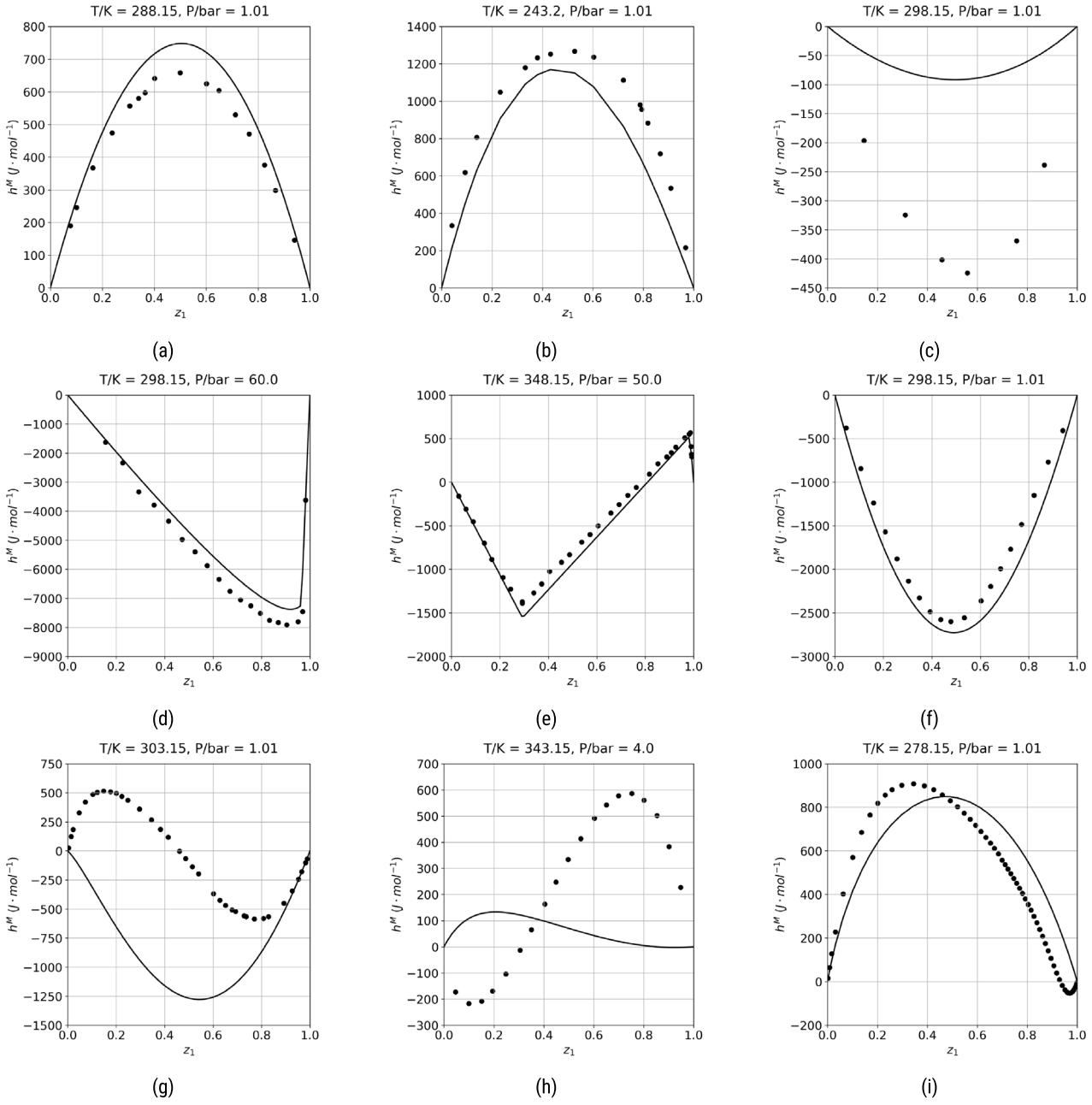

Fig. 14. Enthalpy changes of mixing for nine binary systems calculated with openCOSMO-RS-Phi: •: $h^{M,exp}$, —— $h^{M,calc}$; (a) 1,2-dichloroethane (1) - carbon tetrachloride (BAC_1), (b) acetone (1) - n-hexane (BAC_2), (c) chloroform (1) - benzene (BAC_3), (d) carbon dioxide (1) - dimethyl carbonate (BAC_4), (e) ethane (1) - 1-propanol (BAC_5), (f) ethyl acetate (1) - 1,1,2,2-tetrachloroethane (BAC_6), (g) ethanol (1) - chloroform (BAC_7), (h) acetone (1) - water (BAC_8) and (i) water (1) - acetonitrile (BAC_9).

openCOSMO-RS. In the present implementation, temperature dependence is included only in the dispersion contribution to the free interaction energy, while the misfit and hydrogen-bonding terms are assumed to be temperature independent. Earlier versions of openCOSMO-RS included a temperature-dependent hydrogen-bonding term; however, its omission was found to improve the correlation of water vapor pressures.

One might therefore expect reduced accuracy for systems with significant association effects, such as BAC_5 and BAC_7–BAC_9, where hydrogen bonding plays a dominant role. However, no consistent deterioration is observed. While h^M predictions for BAC_7–BAC_9 are slightly below average, the corresponding c_P^M results for these groups remain above average.

Fig. 14 illustrates the performance for enthalpy-of-mixing data using nine representative binary mixtures, while three examples for the calculation of heat capacities of mixing are presented in Fig. 15.

4. Conclusions

In the present work, openCOSMO-RS-Phi was introduced the first open-source implementation of the full COSMO-RS-based equation of state COSMO-SAC-Phi and its performance was systematically investigated for pure components and binary mixtures. To this end, the model was compared with other equations of state with respect to the reproduction of vapor pressures and molar volumes for up to

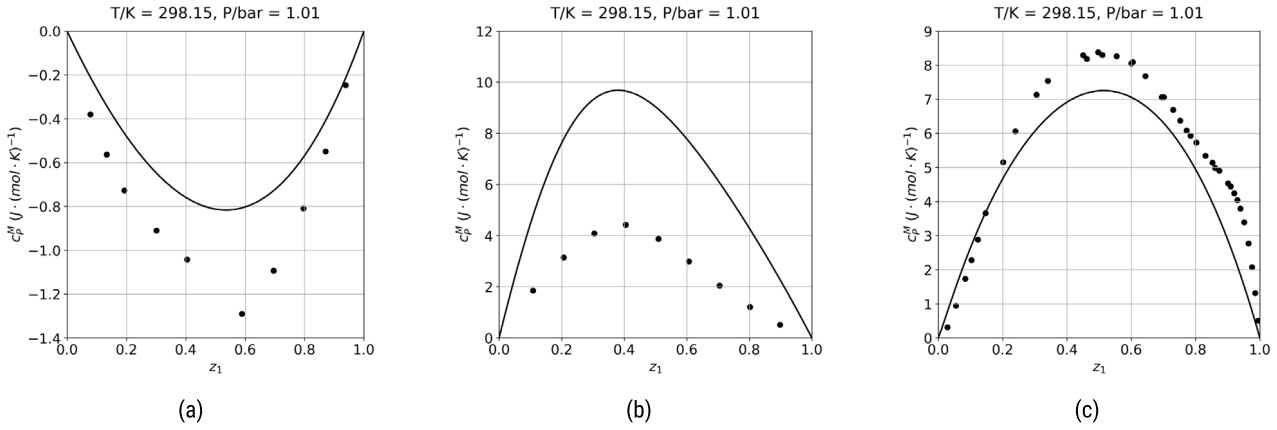

Fig. 15. Isobaric heat capacities of mixing for three binary mixtures calculated with openCOSMO-RS-Phi: •: $c_P^{M,exp}$, —— $c_P^{M,calc}$; (a) cyclohexane (1) - monochlorobenzene (BAC_1), (b) methyl acetate (1) - chloroform (BAC_6), (c) water (1) - acetonitrile (BAC_9).

1800 components. In addition, its performance was evaluated using the binary-mixture database proposed by Jaubert et al..

For pure components, openCOSMO-RS-Phi shows vapor-pressure predictions that are competitive with those of established EoS. For molar volumes, it performs better than several classical equations of state, including SRK, RK, PR, and especially tc-PR. Beyond the numerical results, a central outcome of this work is the availability of an extensive open-source parameter set and a transparent implementation that makes this type of predictive EoS accessible to the wider scientific community.

Based on the systematic analysis of phase-equilibrium and energetic data for binary mixtures, the fully predictive model achieves an overall score of 7.3/20. While quantitative improvements remain possible, openCOSMO-RS-Phi is already capable of qualitatively reproducing a broad range of thermodynamic behavior, including vapor–liquid equilibria, liquid–liquid equilibria, three-phase behavior, azeotropes, and derivative properties such as enthalpies and heat capacities of mixing, without the use of binary interaction parameters. Particularly encouraging is its strong performance for energetic properties and for several classes of associating mixtures.

Overall, this work establishes openCOSMO-RS-Phi as a fully open-source and scientifically useful predictive equation of state based on COSMO-RS theory. Future improvements may include a better repulsive term, refined treatments of misfit, hydrogen-bonding, and dispersion contributions and, where desired, the introduction of binary interaction parameters for more correlative applications.

CRediT authorship contribution statement

Jan Markgraf: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **D.G. Lisboa Girardi:** Writing – original draft, Validation, Methodology, Formal analysis. **Irina Smirnova:** Writing – review & editing, Supervision, Resources, Conceptualization. **Simon Müller:** Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: D. G. Lisboa Girardi reports financial support was provided by German Research Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funded by the Deutsche Forschungsgemeinschaft (DFG) - Project number 503327656.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.fluid.2026.114820.

Data availability

The model implementation, parameter sets, and data generated or analyzed during this study are made openly available in the project repository https://github.com/TUHH-TVT/openCOSMO-RS-Phi_cpp and/or Supporting Information associated with this article.

References

[1] S.-T. Lin, Thermodynamic equations of state from molecular solvation, Fluid Phase Equilib. 245 (2) (2006) 185–192, http://dx.doi.org/10.1016/j.fluid.2006.04.013.

[2] R. Privat, J.-N. Jaubert, The state of the art of cubic equations of state with temperature-dependent binary interaction coefficients: From correlation to prediction, Fluid Phase Equilib. 567 (2023) 113697, http://dx.doi.org/10.1016/j.fluid.2022.113697.

[3] F. Jirasek, R. Bamler, S. Fellenz, M. Bortz, M. Kloft, S. Mandt, H. Hasse, Making thermodynamic models of mixtures predictive by machine learning: matrix completion of pair interactions, Chem. Sci. 13 (2022) 4854–4862, http://dx.doi.org/10.1039/d1sc07210b.

[4] G.M. Kontogeorgis, P. Coutsikos, Thirty years with EoS/g ^{E} models—What have we learned? Ind. Eng. Chem. Res. 51 (11) (2012) 4119–4142, http://dx.doi.org/10.1021/ie2015119.

[5] J.S. Lopez-Echeverry, S. Reif-Acherman, E. Araujo-Lopez, Peng-Robinson equation of state: 40 years through cubics, Fluid Phase Equilib. 447 (2017) 39–71, http://dx.doi.org/10.1016/j.fluid.2017.05.007.

[6] M.-J. Huron, J. Vidal, New mixing rules in simple equations of state for representing vapour-liquid equilibria of strongly non-ideal mixtures, Fluid Phase Equilib. 3 (4) (1979) 255–271, http://dx.doi.org/10.1016/0378-3812(79)80001-1.

[7] M.L. Michelsen, A modified Huron-Vidal mixing rule for cubic equations of state, Fluid Phase Equilib. 60 (1–2) (1990) 213–219, http://dx.doi.org/10.1016/0378-3812(90)85053-D.

[8] D.S.H. Wong, S.I. Sandler, A theoretically correct mixing rule for cubic equations of state, AIChE J. 38 (5) (1992) 671–680, http://dx.doi.org/10.1002/aic.690380505.

[9] E. Voutsas, K. Magoulas, D. Tassios, Universal mixing rule for cubic equations of state applicable to symmetric and asymmetric systems: Results with the Peng-Robinson equation of state, Ind. Eng. Chem. Res. 43 (19) (2004) 6238–6246, http://dx.doi.org/10.1021/ie049580p.

[10] R. Privat, J.-N. Jaubert, G.M. Kontogeorgis, Let us rethink advanced mixing rules for cubic equations of state, Fluid Phase Equilib. 596 (2025-09-01) 114455, http://dx.doi.org/10.1016/j.fluid.2025.114455.
[11] D.-Y. Peng, D.B. Robinson, A new two-constant equation of state, Ind. Eng. Chem. Fundam. 15 (1) (1976) 59–64, http://dx.doi.org/10.1021/i160057a011.
[12] G. Soave, Equilibrium constants from a modified Redlich-Kwong equation of state, Chem. Eng. Sci. 27 (6) (1972) 1197–1203, http://dx.doi.org/10.1016/0009-2509(72)80096-4.
[13] A. Fredenslund, R.L. Jones, J.M. Prausnitz, Group-contribution estimation of activity coefficients in nonideal liquid mixtures, AIChE J. 21 (6) (1975) 1086–1099, http://dx.doi.org/10.1002/aic.690210607.
[14] J. Gmehling, J. Li, M. Schiller, A modified UNIFAC model. 2. Present parameter matrix and results for different thermodynamic properties, Ind. Eng. Chem. Res. 32 (1) (1993) 178–193, http://dx.doi.org/10.1021/ie00013a024.
[15] W. Yan, M. Topphoff, C. Rose, J. Gmehling, Prediction of vapor–liquid equilibria in mixed-solvent electrolyte systems using the group contribution concept, Fluid Phase Equilib. 162 (1–2) (1999) 97–113, http://dx.doi.org/10.1016/S0378-3812(99)00201-0.
[16] K. Kojima, K. Tochigi, Prediction of Vapor-Liquid Equilibria by the ASOG Method, first ed., Kodansha Ltd., Tokyo, ISBN: 0-444-99773-3, 1979.
[17] T. Holderbaum, J. Gmehling, PSRK: A group contribution equation of state based on UNIFAC, Fluid Phase Equilib. 70 (2–3) (1991) 251–265, http://dx.doi.org/10.1016/0378-3812(91)85038-V.
[18] C. Boukouvalas, N. Spiliotis, P. Coutsikos, N. Tzouvaras, D. Tassios, Prediction of vapor-liquid equilibrium with the LCVM model: a linear combination of the Vidal and michelsen mixing rules coupled with the original UNIF, Fluid Phase Equilib. 92 (1994) 75–106, http://dx.doi.org/10.1016/0378-3812(94)80043-X.
[19] J. Ahlers, J. Gmehling, Development of an universal group contribution equation of state I. prediction of liquid densities for pure compounds with a volume translated Peng–Robinson equation of state, Fluid Phase Equilib. 191 (1–2) (2001) 177–188, http://dx.doi.org/10.1016/S0378-3812(01)00626-4.
[20] J. Ahlers, J. Gmehling, Development of a universal group contribution equation of state. 2. Prediction of vapor-liquid equilibria for asymmetric systems, Ind. Eng. Chem. Res. 41 (14) (2002) 3489–3498, http://dx.doi.org/10.1021/ie020047o.
[21] J. Ahlers, J. Gmehling, Development of a universal group contribution equation of state III. Prediction of vapor-liquid equilibria, excess enthalpies, and activity coefficients at infinite dilution with the VTPR model, Ind. Eng. Chem. Res. 41 (23) (2002) 5890–5899, http://dx.doi.org/10.1021/ie0203734.
[22] L.-S. Wang, J. Ahlers, J. Gmehling, Development of a universal group contribution equation of state. 4. Prediction of vapor-liquid equilibria of polymer solutions with the volume translated group contribution equation of state, Ind. Eng. Chem. Res. 42 (24) (2003) 6205–6211, http://dx.doi.org/10.1021/ie0210356.
[23] J. Ahlers, T. Yamaguchi, J. Gmehling, Development of a universal group contribution equation of state. 5. Prediction of the solubility of high-boiling compounds in supercritical gases with the group contribution equation of state volume-translated Peng-Robinson, Ind. Eng. Chem. Res. 43 (20) (2004) 6569–6576, http://dx.doi.org/10.1021/ie040037i.
[24] E. Voutsas, V. Louli, C. Boukouvalas, K. Magoulas, D. Tassios, Thermodynamic property calculations with the universal mixing rule for eos/GE models: Results with the Peng–Robinson Eos and a UNIFAC model, Fluid Phase Equilib. 241 (1–2) (2006) 216–228, http://dx.doi.org/10.1016/j.fluid.2005.12.028.
[25] S. Dahl, M.L. Michelsen, High-pressure vapor-liquid equilibrium with a UNIFAC-based equation of state, AIChE J. 36 (12) (1990) 1829–1836, http://dx.doi.org/10.1002/aic.690361207.
[26] K. Tochigi, Prediction of high-pressure vapor-liquid equilibria using ASOG, Fluid Phase Equilib. 104 (1995) 253–260, http://dx.doi.org/10.1016/0378-3812(94)02652-H.
[27] J. Li, M. Topphoff, K. Fischer, J. Gmehling, Prediction of gas solubilities in aqueous electrolyte systems using the predictive Soave-Redlich-Kwong model, Ind. Eng. Chem. Res. 40 (16) (2001) 3703–3710, http://dx.doi.org/10.1021/ie0100535.
[28] J. Gross, G. Sadowski, Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules, Ind. Eng. Chem. Res. 40 (4) (2001) 1244–1260, http://dx.doi.org/10.1021/ie0003887.
[29] J. Gross, G. Sadowski, Application of the perturbed-chain SAFT equation of state to associating systems, Ind. Eng. Chem. Res. 41 (22) (2002) 5510–5515, http://dx.doi.org/10.1021/ie010954d.
[30] P.J. Walker, Toward advanced, predictive mixing rules in SAFT equations of state, Ind. Eng. Chem. Res. 61 (49) (2022) 18165–18175, http://dx.doi.org/10.1021/acs.iecr.2c03464.
[31] G.M. Kontogeorgis, E.C. Voutsas, I.V. Yakoumis, D.P. Tassios, An equation of state for associating fluids, Ind. Eng. Chem. Res. 35 (11) (1996) 4310–4318, http://dx.doi.org/10.1021/ie9600203.
[32] G. Tasios, V. Louli, S. Skouras, E. Solbraa, E. Voutsas, Development of a new group contribution equation of state for associating compounds, Fluid Phase Equilib. 571 (2023) 113824, http://dx.doi.org/10.1016/j.fluid.2023.113824.
[33] C. Panayiotou, D. Aslanidou, Partial solvation parameters and the equation-of-state approach, Fluid Phase Equilib. 406 (2015) 101–115, http://dx.doi.org/10.1016/j.fluid.2015.08.004.
[34] S. Mastrogeorgopoulos, V. Hatzimanikatis, C. Panayiotou, Toward a simple predictive molecular thermodynamic model for bulk phases and interfaces, Ind. Eng. Chem. Res. 56 (38) (2017) 10900–10910, http://dx.doi.org/10.1021/acs.iecr.7b02286.
[35] W. Chapman, K. Gubbins, G. Jackson, M. Radosz, SAFT: Equation-of-state solution model for associating fluids, Fluid Phase Equilib. 52 (1989) 31–38, http://dx.doi.org/10.1016/0378-3812(89)80308-5.
[36] W.G. Chapman, K.E. Gubbins, G. Jackson, M. Radosz, New reference equation of state for associating liquids, Ind. Eng. Chem. Res. 29 (8) (1990) 1709–1721, http://dx.doi.org/10.1021/ie00104a021.
[37] F. Shaahmadi, S.A. Smith, C.E. Schwarz, A.J. Burger, J.T. Cripwell, Group-contribution SAFT equations of state: A review, Fluid Phase Equilib. 565 (2023) 113674, http://dx.doi.org/10.1016/j.fluid.2022.113674.
[38] K. Zhu, E.A. Müller, Generating a machine-learned equation of state for fluid properties, J. Phys. Chem. B 124 (39) (2020) 8628–8639, http://dx.doi.org/10.1021/acs.jpcb.0c05806.
[39] H. Matsukawa, M. Kitahara, K. Otake, Estimation of pure component parameters of PC-SAFT EoS by an artificial neural network based on a group contribution method, Fluid Phase Equilib. 548 (2021) 113179, http://dx.doi.org/10.1016/j.fluid.2021.113179.
[40] A. Abdallah El Hadj, M. Laidi, S. Hanini, AI-PCSAFT approach: New high predictive method for estimating PC-SAFT pure component properties and phase equilibria parameters, Fluid Phase Equilib. 555 (2022) 113297, http://dx.doi.org/10.1016/j.fluid.2021.113297.
[41] J. Habicht, C. Brandenbusch, G. Sadowski, Predicting PC-SAFT pure-component parameters by machine learning using a molecular fingerprint as key input, Fluid Phase Equilib. 565 (2023) 113657, http://dx.doi.org/10.1016/j.fluid.2022.113657.
[42] B. Winter, P. Rehner, T. Esper, J. Schilling, A. Bardow, Understanding the language of molecules: Predicting pure component parameters for the PC-SAFT equation of state from SMILES, 2023, http://dx.doi.org/10.48550/ARXIV.2309.12404, Version Number: 1.
[43] K.C. Felton, L. Raßpe-Lange, J.G. Rittig, K. Leonhard, A. Mitsos, J. Meyer-Kirschner, C. Knösche, A.A. Lapkin, ML-SAFT: A machine learning framework for PCP-SAFT parameter prediction, Chem. Eng. J. 492 (2024) 151999, http://dx.doi.org/10.1016/j.cej.2024.151999.
[44] J.-N. Jaubert, F. Mutelet, VLE predictions with the Peng–Robinson equation of state and temperature dependent kij calculated through a group contribution method, Fluid Phase Equilib. 224 (2) (2004) 285–304, http://dx.doi.org/10.1016/j.fluid.2004.06.059.
[45] X. Xu, S. Lasala, R. Privat, J.-N. Jaubert, E -PPR78: A proper cubic EoS for modelling fluids involved in the design and operation of carbon dioxide capture and storage (CCS) processes, Int. J. Greenh. Gas Control. 56 (2017) 126–154, http://dx.doi.org/10.1016/j.ijggc.2016.11.015.
[46] J.-N. Jaubert, J.-W. Qian, S. Lasala, R. Privat, The impressive impact of including enthalpy and heat capacity of mixing data when parameterising equations of state. Application to the development of the E-PPR78 (enhanced-predictive-Peng-Robinson-78) model, Fluid Phase Equilib. 560 (2022) 113456, http://dx.doi.org/10.1016/j.fluid.2022.113456.
[47] A.M. Abudour, S.A. Mohammad, R.L. Robinson Jr., K.A. Gasem, Generalized binary interaction parameters for the Peng–Robinson equation of state, Fluid Phase Equilib. 383 (2014) 156–173, http://dx.doi.org/10.1016/j.fluid.2014.10.006.
[48] A.M. Abudour, S.A. Mohammad, R.L. Robinson, K.A. Gasem, Predicting PR EOS binary interaction parameter using readily available molecular properties, Fluid Phase Equilib. 434 (2017) 130–140, http://dx.doi.org/10.1016/j.fluid.2016.11.019.
[49] M. Stavrou, A. Bardow, J. Gross, Estimation of the binary interaction parameter k of the PC-SAFT equation of state based on pure component parameters using a QSPR method, Fluid Phase Equilib. 416 (2016) 138–149, http://dx.doi.org/10.1016/j.fluid.2015.12.016.
[50] F. Abbasi, Z. Abbasi, R. Bozorgmehry Boozarjomehry, Estimation of PC-SAFT binary interaction coefficient by artificial neural network for multicomponent phase equilibrium calculations, Fluid Phase Equilib. 510 (2020) 112486, http://dx.doi.org/10.1016/j.fluid.2020.112486.
[51] A. Klamt, Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena, J. Phys. Chem. 99 (7) (1995) 2224–2235, http://dx.doi.org/10.1021/j100007a062.
[52] A. Klamt, V. Jonas, T. Bürger, J.C.W. Lohrenz, Refinement and parametrization of COSMO-RS, J. Phys. Chem. A 102 (26) (1998) 5074–5085, http://dx.doi.org/10.1021/jp980017s.
[53] A. Klamt, F. Eckert, COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids, Fluid Phase Equilib. 172 (1) (2000) 43–72, http://dx.doi.org/10.1016/S0378-3812(00)00357-5.
[54] H. Grensemann, J. Gmehling, Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods, Ind. Eng. Chem. Res. 44 (5) (2005) 1610–1624, http://dx.doi.org/10.1021/ie049139z.
[55] S.-T. Lin, S.I. Sandler, A priori phase equilibrium prediction from a segment contribution solvation model, Ind. Eng. Chem. Res. 41 (5) (2002) 899–913, http://dx.doi.org/10.1021/ie001047w.

[56] Y. Shimoyama, Y. Iwai, Development of activity coefficient model based on COSMO method for prediction of solubilities of solid solutes in supercritical carbon dioxide, J. Supercrit. Fluids 50 (3) (2009) 210–217, http://dx.doi.org/10.1016/j.supflu.2009.06.004.

[57] C.-M. Hsieh, S.-T. Lin, J. Vrabec, Considering the dispersive interactions in the COSMO-SAC model for more accurate predictions of fluid phase behavior, Fluid Phase Equilib. 367 (2014) 109–116, http://dx.doi.org/10.1016/j.fluid.2014.01.032.

[58] R.D.P. Soares, R.P. Gerber, L.F.K. Possani, P.B. Staudt, Functional-segment activity coefficient model. 2. Associating mixtures, Ind. Eng. Chem. Res. 52 (32) (2013) 11172–11181, http://dx.doi.org/10.1021/ie4013979.

[59] R.D.P. Soares, R.P. Gerber, Functional-segment activity coefficient model. 1. Model formulation, Ind. Eng. Chem. Res. 52 (32) (2013) 11159–11171, http://dx.doi.org/10.1021/ie400170a.

[60] D. Constantinescu, A. Klamt, D. Geană, Vapor–liquid equilibrium prediction at high pressures using activity coefficients at infinite dilution from COSMO-type methods, Fluid Phase Equilib. 231 (2) (2005) 231–238, http://dx.doi.org/10.1016/j.fluid.2005.01.014.

[61] Y. Shimoyama, Y. Iwai, S. Takada, Y. Arai, T. Tsuji, T. Hiaki, Prediction of phase equilibria for mixtures containing water, hydrocarbons and alcohols at high temperatures and pressures by cubic equation of state with GE type mixing rule based on COSMO-RS, Fluid Phase Equilib. 243 (1–2) (2006) 183–192, http://dx.doi.org/10.1016/j.fluid.2006.03.007.

[62] M.-T. Lee, S.-T. Lin, Prediction of mixture vapor–liquid equilibrium from the combined use of Peng–Robinson equation of state and COSMO-SAC activity coefficient model through the wong–sandler mixing rule, Fluid Phase Equilib. 254 (1–2) (2007) 28–34, http://dx.doi.org/10.1016/j.fluid.2007.02.012.

[63] Y. Shimoyama, T. Abeta, Y. Iwai, Prediction of vapor–liquid equilibria for supercritical alcohol+fatty acid ester systems by SRK equation of state with Wong–Sandler mixing rule based on COSMO theory, J. Supercrit. Fluids 46 (1) (2008) 4–9, http://dx.doi.org/10.1016/j.supflu.2008.02.013.

[64] K. Leonhard, J. Veverka, K. Lucas, A comparison of mixing rules for the combination of COSMO-RS and the Peng–Robinson equation of state, Fluid Phase Equilib. 275 (2) (2009) 105–115, http://dx.doi.org/10.1016/j.fluid.2008.09.016.

[65] P.B. Staudt, R.D.P. Soares, A self-consistent Gibbs excess mixing rule for cubic equations of state, Fluid Phase Equilib. 334 (2012) 76–88, http://dx.doi.org/10.1016/j.fluid.2012.06.029.

[66] C.-Y. Chen, L.-H. Wang, C.-M. Hsieh, S.-T. Lin, Prediction of solid-liquid-gas equilibrium for binary mixtures of carbon dioxide + organic compounds from approaches based on the COSMO-SAC model, J. Supercrit. Fluids 133 (2018) 318–329, http://dx.doi.org/10.1016/j.supflu.2017.08.008.

[67] L.-H. Wang, C.-M. Hsieh, S.-T. Lin, Prediction of gas and liquid solubility in organic polymers based on the PR+COSMOSAC equation of State, Ind. Eng. Chem. Res. 57 (31) (2018) 10628–10639, http://dx.doi.org/10.1021/acs.iecr.8b01780.

[68] D. Mambo-Lomba, P. Paricaud, Predictions of thermodynamic properties and phase equilibria of refrigerant systems with COSMO approaches, Int. J. Refrig. 124 (2021) 52–63, http://dx.doi.org/10.1016/j.ijrefrig.2020.11.005.

[69] J. Wang, T. Sa, L. Zhu, Vapor-liquid equilibrium predictions of refrigerant systems using COSMO based Gex-EoS methods, Fluid Phase Equilib. 563 (2023) 113584, http://dx.doi.org/10.1016/j.fluid.2022.113584.

[70] F.C. Paes, R. Privat, J.-N. Jaubert, B. Sirjean, Prediction of solvation energies at infinite dilution by the tc-PR cubic equation of state with advanced mixing rule based on COSMO-RS as gE model, J. Mol. Liq. 386 (2023) 122480, http://dx.doi.org/10.1016/j.molliq.2023.122480.

[71] F. Paes, R. Privat, J.-N. Jaubert, B. Sirjean, Predicting solvation energies of free radicals and their mixtures: A robust approach coupling the Peng-Robinson and COSMO-RS models, J. Mol. Liq. 401 (2024) 124641, http://dx.doi.org/10.1016/j.molliq.2024.124641.

[72] N. Prinos, E. Voutsas, Phase equilibrium calculations at low and high pressures with a modified COSMO-SAC model, Fluid Phase Equilib. 589 (2025) 114277, http://dx.doi.org/10.1016/j.fluid.2024.114277.

[73] C. Song, R. Shariyati, I. Alkhrsan, Prediction of thermodynamic properties of ionic liquids using the PC-SAFT EoS coupled with COSMO-RS model, Chem. Eng. Res. Des. 213 (2025) 1–10, http://dx.doi.org/10.1016/j.cherd.2024.11.030.

[74] C. Panayiotou, Solvation thermodynamics and non-randomness. Part I: Self-solvation, J. Chem. Eng. Data 55 (12) (2010) 5453–5464, http://dx.doi.org/10.1021/je100575q.

[75] C. Panayiotou, Toward a COSMO equation-of-state model of fluids and their mixtures, Pure Appl. Chem. 83 (6) (2011) 1221–1242, http://dx.doi.org/10.1351/PAC-CON-10-08-14.

[76] C.T. Costa, F.W. Tavares, A.R. Secchi, Equation of state based on the hole-lattice theory and surface-charge density (COSMO): Part b – vapor–liquid equilibrium for mixtures, Fluid Phase Equilib. 419 (2016) 1–10, http://dx.doi.org/10.1016/j.fluid.2016.03.005.

[77] C.T. Costa, F.W. Tavares, A.R. Secchi, Equation of state based on the hole-lattice theory and surface-charge density (COSMO): Part A – pure compounds, Fluid Phase Equilib. 409 (2016) 472–481, http://dx.doi.org/10.1016/j.fluid.2015.11.010.

[78] R. De P. Soares, L.F. Baladão, P.B. Staudt, A pairwise surface contact equation of state: COSMO-SAC-phi, Fluid Phase Equilib. 488 (2019) 13–26, http://dx.doi.org/10.1016/j.fluid.2019.01.015.

[79] C. Panayiotou, W. Acree, I. Zuburtikudis, COSMO-RS and LSER models of solution thermodynamics: Towards a COSMO-LSER equation of state model of fluids, J. Mol. Liq. 390 (2023) 122992, http://dx.doi.org/10.1016/j.molliq.2023.122992.

[80] T. Gerlach, S. Müller, A.G. De Castilla, I. Smirnova, An open source COSMO-RS implementation and parameterization supporting the efficient implementation of multiple segment descriptors, Fluid Phase Equilib. 560 (2022) 113472, http://dx.doi.org/10.1016/j.fluid.2022.113472.

[81] L.P. Zini, P.B. Staudt, R.D.P. Soares, An improved dispersive contribution for the COSMO-SAC-Phi equation of state, Fluid Phase Equilib. 534 (2021) 112942, http://dx.doi.org/10.1016/j.fluid.2021.112942.

[82] L.F. Baladao, P.B. Staudt, R. de P. Soares, Functional-Segment activity coefficient equation of state: F-SAC-Phi, Ind. Eng. Chem. Res. 58 (2019) 16934–16944, http://dx.doi.org/10.1021/acs.iecr.9b02190.

[83] Dassault Systèmes, BIOVIA COSMOtherm 2025 - Reference Manual, Vélizy-Villacoublay, 2024, Release 2025. Available at https://www.3ds.com/products/biovia/cosmo-rs/cosmotherm.

[84] J.M. Prausnitz, R.N. Lichtenthaler, E.G.d. Azevedo, Molecular Thermodynamics of Fluid-Phase Equilibria, third ed., in: Prentice-Hall international series in the physical and chemical engineering sciences, Prentice Hall PTR, Upper Saddle River, N.J, ISBN: 978-0-13-977745-5, 1999.

[85] B.E. Poling, J.M. Prausnitz, J.P. O'Connell, The Properties of Gases and Liquids, fifth ed., McGraw-Hill, New York, ISBN: 978-0-07-011682-5, 2001.

[86] M.L. Michelsen, J.M. Mollerup, Thermodynamic Models: Fundamentals & Computational Aspects, second ed, Tie-Line Publications, Holte, ISBN: 978-87-989961-3-2, 2007.

[87] A. Klamt, G.J.P. Krooshof, R. Taylor, COSMOSPACE: Alternative to conventional activity-coefficient models, AIChE J. 48 (10) (2002) 2332–2349, http://dx.doi.org/10.1002/aic.690481023.

[88] S. Müller, T. Nevolianis, M. Garcia-Ratés, C. Riplinger, K. Leonhard, I. Smirnova, Predicting solvation free energies for neutral molecules in any solvent with opencosmo-RS, Fluid Phase Equilib. 589 (2025) 114250, http://dx.doi.org/10.1016/j.fluid.2024.114250.

[89] J.L.O. W.Vincent Wilding, Dippr® project 801 evaluated process design data, Fluid Phase Equilib. 150–151 (1998) 413–420, http://dx.doi.org/10.1016/S0378-3812(98)00341-0.

[90] A. Piña-Martinez, R. Privat, J.-N. Jaubert, 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. 68, 2 (2021) e17518, http://dx.doi.org/10.1002/aic.17518.

[91] N. Ramírez-Vélez, R. Privat, A. Piña-Martinez, J.-N. Jaubert, Assessing the performance of non-associating SAFT-type equations of state to reproduce vapor pressure, liquid density, enthalpy of vaporization, and liquid heat capacity data of 1800 pure fluids, AIChE J. 68 (7) (2022) e17722, http://dx.doi.org/10.1002/aic.17722.

[92] J.A. Nelder, R. Mead, A simplex method for function minimization, Comput. J. 7 (4) (1965) 308–313, http://dx.doi.org/10.1093/comjnl/7.4.308.

[93] R. Storn, K. Price, Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces, J. Global Optim. 11 (1997) 341–359, http://dx.doi.org/10.1023/A:1008202821328.

[94] J.-N. Jaubert, Y. Le Guennec, A. Piña-Martinez, N. Ramirez-Velez, S. Lasala, B. Schmid, I.K. Nikolaidis, I.G. Economou, R. Privat, Benchmark database containing binary-system-high-quality-certified data for cross-comparing thermodynamic models and assessing their accuracy, Ind. Eng. Chem. Res. 59 (33) (2020) 14981–15027, http://dx.doi.org/10.1021/acs.iecr.0c01734.

[95] A. Piña-Martinez, R. Privat, I.K. Nikolaidis, I.G. Economou, J.-N. Jaubert, What is the optimal activity coefficient model to be combined with the translated–consistent Peng–Robinson equation of state through advanced mixing rules? Cross-comparison and grading of the wilson, UNIQUAC, and NRTL a E models against a benchmark database involving 200 binary systems, Ind. Eng. Chem. Res. 60 (47) (2021) 17228–17247, http://dx.doi.org/10.1021/acs.iecr.1c03003.

[96] F. Paes, B. Sirjean, J.-N. Jaubert, R. Privat, Using COSMO-SAC models in the mixing rules of the translated-consistent peng-robinson equation-of-state: a rigorous benchmarking across a broad and diverse database, Chem. Eng. Sci. (2026) 123300, http://dx.doi.org/10.1016/j.ces.2026.123300.

[97] I.K. Nikolaidis, R. Privat, J.-N. Jaubert, I.G. Economou, 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. 60 (24) (2021) 8935–8946, http://dx.doi.org/10.1021/acs.iecr.1c01234.

[98] I.K. Nikolaidis, R. Privat, J.-N. Jaubert, I.G. Economou, The effect of binary interaction parameters and induced association on the performance of the PC-SAFT equation of state: validation against a benchmark database of high-quality binary-system data, J. Chem. Eng. Data 69 (2) (2023) 320–337, http://dx.doi.org/10.1021/acs.jced.2c00781.

[99] J.R. Elliott, C.T. Lira, Introductory Chemical Engineering Thermodynamics, second ed., Prentice Hall, Upper Saddle River, NJ, ISBN: 978-0-13-606854-9, 2012.

[100] M.O. Ohanamah, D.W. Thompson, Computation of multicomponent phase equilibria—Part ii. liquid-liquid and solid-liquid equilibria, Comput. Chem. Eng. 8 (3–4) (1984) 157–162, http://dx.doi.org/10.1016/0098-1354(84)87002-7.

[101] R. Privat, J.-N. Jaubert, Discussion around the paradigm of ideal mixtures with emphasis on the definition of the property changes on mixing, Chem. Eng. Sci. 82 (2012) 319–333, http://dx.doi.org/10.1016/j.ces.2012.07.030.

[102] J.-W. Qian, R. Privat, J.-N. Jaubert, P. Duchet-Suchaux, Enthalpy and heat capacity changes on mixing: fundamental aspects and prediction by means of the PPR78 cubic equation of state, Energy Fuels 27 (11) (2013) 7150–7178, http://dx.doi.org/10.1021/ef401605c.

[103] W. Yan, A robust and efficient solution for COSMO-based activity coefficient models, Chem. Eng. Sci. 300 (2024) 120652, http://dx.doi.org/10.1016/j.ces.2024.120652.

[104] N. Ramírez-Vélez, A. Pina-Martinez, J.-N. Jaubert, R. Privat, 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 65 (12) (2020) 5920–5932, http://dx.doi.org/10.1021/acs.jced.0c00792.

[105] F. Paes, R. Privat, R. Bounaceur, J.-N. Jaubert, Prediction of pure-component thermodynamic properties using the Peng–Robinson equation of state with machine learning-derived input parameters, Ind. Eng. Chem. Res. (2026) http://dx.doi.org/10.1021/acs.iecr.6c00300.

[106] Ø. Wilhelmsen, A. Aasen, G. Skaugen, P. Aursand, A. Austegard, E. Aursand, M.A. Gjennestad, H. Lund, G. Linga, M. Hammer, Thermodynamic modeling with equations of state: present challenges with established methods, Ind. Eng. Chem. Res. 56 (13) (2017) 3503–3515, http://dx.doi.org/10.1021/acs.iecr.7b00317.

[107] M. Cismondi, E.A. Brignole, J. Mollerup, Rescaling of three-parameter equations of state: PC-SAFT and SPHCT, Fluid Phase Equilib. 234 (1–2) (2005) 108–121, http://dx.doi.org/10.1016/j.fluid.2005.06.002.

[108] A.M. Palma, M.B. Oliveira, A.J. Queimada, J.A. Coutinho, Re-evaluating the CPA EoS for improving critical points and derivative properties description, Fluid Phase Equilib. 436 (2017) 85–97, http://dx.doi.org/10.1016/j.fluid.2017.01.002.

[109] E.C. Voutsas, G.D. Pappa, K. Magoulas, D.P. Tassios, Vapor liquid equilibrium modeling of alkane systems with equations of state:“Simplicity versus complexity”, Fluid Phase Equilib. 240 (2) (2006) 127–139, http://dx.doi.org/10.1016/j.fluid.2005.12.010.

[110] K. Li, Z. Ye, X. Liang, Comparative study of phase equilibrium modeling with cubic and association equations of state, J. Chem. Eng. Data 69 (12) (2024) 4261–4279, http://dx.doi.org/10.1021/acs.jced.4c00204.