Chemical Engineering Science 319 (2026) 122170

Contents lists available at ScienceDirect

Chemical Engineering Science

journal homepage: www.elsevier.com/locate/ces

A comprehensive approach to incorporating intermolecular dispersion into the openCOSMO-RS model. Part 2: Atomic polarizabilities

Daria Grigorash [a,b], Simon Müller [c], Esther Heid [d], Frank Neese [e], Dimitrios G. Liakos [f], Christoph Riplinger [e], Miquel Garcia-Ratés [e], Patrice Paricaud [g], Erling H. Stenby [a,b], Irina Smirnova [c], Wei Yan [a,b,*]

[a] Department of Chemistry, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
[b] Center for Energy Resources Engineering, Technical University of Denmark, Kgs. Lyngby, 2800, Denmark
[c] Institute of Thermal Separation Processes, Hamburg University of Technology, Hamburg, 21073, Germany
[d] Institute of Material Chemistry, TU Wien, Vienna, 1060, Austria
[e] FAccTs GmbH, Cologne, 50677, Germany
[f] Max-Planck-Institut für Kohlenforschung, Muelheim an der Ruhr, 45470, Germany
[g] UCP, ENSTA Paris, Institut Polytechnique de Paris, Palaiseau, 91762, France

ARTICLE INFO

Keywords:
COSMO-RS
Dispersion
Parametrization
Polarizability

ABSTRACT

OpenCOSMO-RS is an open-source predictive thermodynamic model that can be applied to a broad range of systems in various chemical and biochemical engineering domains. This study focuses on improving openCOSMO-RS by introducing a new dispersion term based on atomic polarizabilities. We evaluate different methods for processing polarizability data, including scaling and combining it to compute segment-segment dispersion interaction energies, with a focus on halocarbon systems. The results demonstrate that the modified model outperforms our previous method developed in the first part of this work Grigorash et al. (2024), while at the same time requiring fewer adjustable parameters. The approach was applied to a broad dataset of over 50,000 data points, consistently increasing the accuracy across a variety of data types. These findings suggest that atomic polarizability is a valuable descriptor for refining dispersion interactions in predictive thermodynamic models.

1. Introduction

Predictive thermodynamic models that integrate molecular-level quantum chemical (QC) information have become essential tools in both academic research and industrial applications. A prominent approach in this domain is the COSMO-RS (COnductor-like Screening Model for Realistic Solvents) model, developed by combining QC calculations of screening surface charge densities with statistical thermodynamics (Klamt, 1995; Klamt et al., 1998). It offers a predictive framework for estimating thermodynamic properties directly from molecular descriptors.

Since the initial development of COSMO-RS, several implementations have been proposed including COSMO-SAC (Segment Activity Coefficient) (Lin and Sandler, 2002), COSMO-RS(ol) (Grensemann and Gmehling, 2005) and openCOSMO-RS (Gerlach et al., 2022). COSMO-SAC is the first open-source implementation closely based on the original COSMO-RS, with minor adjustments in the derivation and interpretation of the model. However, subsequent developments introduced more significant deviations from the original COSMO-RS, including the treatment of hydrogen bonding and dispersive interactions. In addition, these models rely on different quantum chemical packages and, consequently, different general parameters.

Various COSMO-RS implementations have been successfully applied to predicting thermodynamic properties such as vapor-liquid equilibria (VLE) (Klamt and Eckert, 2007; Mambo-Lomba and Paricaud, 2021; Hsieh et al., 2010, 2014; Grensemann and Gmehling, 2005), infinite dilution activity coefficients (IDACs) (Fingerhut et al., 2017; Gerlach et al., 2022), solvation free energies (Müller et al., 2025; Saidi et al., 2020; Paes et al., 2024), partitioning coefficients (Gerlach et al., 2022; Klamt et al., 1998) and phase behavior across diverse chemical families (Jiřiště and Klajmon, 2022; Klajmon, 2022; Peng et al., 2022), including both polar and nonpolar (Klamt et al., 2002; Klamt, 2003; Eckert and Klamt, 2003), and electrolyte systems (Gerlach et al., 2018; Müller et al., 2019; Kröger et al., 2020; Müller et al., 2020; González de Castilla et al., 2021; Ritter et al., 2016; Arrad et al., 2024; González de Castilla et al., 2022).

* Corresponding author.
E-mail address: weya@kemi.dtu.dk (W. Yan).

https://doi.org/10.1016/j.ces.2025.122170
Received 11 February 2025; Received in revised form 19 June 2025; Accepted 2 July 2025
Available online 9 July 2025

Dispersion forces–also known as London dispersion forces–arise from interactions between instantaneous dipoles generated by fluctuations in polarizable electron clouds (Prausnitz et al., 1999). These forces are present in all molecular systems, regardless of polarity, and become particularly important in nonpolar and weakly polar compounds, such as halocarbons. The strength of dispersion interactions is intrinsically linked to the polarizability of the atoms or molecules involved, which measures how easily an electron cloud can be distorted by an external electric field. A more easily distorted (i.e., more polarizable) electron cloud facilitates the formation of instantaneous dipoles, thereby enhancing dispersion interactions. Typically, atomic polarizability increases with atomic size, as the Coulombic attraction between the nucleus and outer electrons weakens. This makes heavier atoms, such as certain halogens, more polarizable than lighter atoms. From a theoretical perspective, dispersion interactions can be quantitatively described by London's dispersion potential, which is proportional to the product of the polarizabilities of the interacting species and inversely proportional to the sixth power of their separation distance (London, 1937). This formalism highlights the central role of polarizability in describing van der Waals interactions and suggests that it can serve as a critical link between quantum mechanical properties and macroscopic thermodynamic behavior. Despite this fundamental importance, existing COSMO-RS implementations often treat dispersion interactions in a highly simplified or empirical manner. Several models (Hsieh et al., 2014; Klamt and Eckert, 2007) use adjustable parameters that are not clearly grounded in quantum mechanical descriptors. As a result, the predictive accuracy of these models may deteriorate in systems where dispersion forces are dominant. To address these limitations, this study introduces a new dispersion term for the openCOSMO-RS framework that explicitly incorporates atomic polarizability as a QC derived descriptor.

In the first part of this work (Grigorash et al., 2024), we developed a dispersion term for the openCOSMO-RS model based on a small set of general, adjustable atomic parameters, correlated with atomic polarizabilities. This led to significant improvements in modeling the phase equilibria of halocarbons. Building on that foundation, the focus of this work is the development of a generalized openCOSMO-RS model that incorporates a dispersion term based on a novel descriptor–atomic polarizability–which, to the best of our knowledge, has not been previously implemented in the COSMO-RS framework. Atomic polarizability tensors were calculated using the latest release of ORCA 6.0 (Neese, 2012; Neese et al., 2020; Garcia-Ratés and Neese, 2020), as detailed in Section 2.1, and subsequently projected onto molecular cavities following the procedure outlined in Section 2.2. The approach for incorporating polarizabilities into the COSMO-RS equations is discussed in Section 2.3, followed by the description of the regression procedure used to fine-tune the modified model to experimental data (Section 2.4). As in the first part of this study, we initially focused on halocarbons 3.1, as these systems are particularly sensitive to dispersive interactions and frequently require a special consideration of the dispersion term for an adequate modeling (Klamt and Eckert, 2007). In Section 3.1.1, we discuss these challenging interactions in depth, focusing primarily on hydrocarbon - fluorocarbon mixtures. To determine the optimal approach for incorporating polarizabilities into the dispersion term using halocarbon data, we evaluate various methodologies in Sections 3.1.2 and 3.1.3. Additionally, we investigate the impact of the F-atom radius used in C-PCM (Conductor-like Polarizable Continuum Model) calculations (Section 3.1.4) and different polarizability combining strategies in Sections 3.1.5 and 3.1.6. Finally, we extend the models that demonstrated the best performance for halocarbon data (Section 3.1.7) to a comprehensive dataset (Section 3.2) of nearly 50,000 data points, covering compounds of diverse chemical nature, which results in the improved general openCOSMO-RS model.

2. Methods

2.1. QC calculations

For QC C-PCM calculations, we used the open-source workflow from the first part of this work and detailed in Gerlach et al. (2022), now with the latest ORCA 6.0 release. The new version includes methods for calculating molecular static polarizability tensors and decomposing them into atomic components. These properties are calculated by calling an additional block for electrical properties, along with the reference method, such as Hartree-Fock or DFT (Density Functional Theory), to solve for the electronic structure.

An example input might contain the following block:

```
\%elprop
  Polar 1
  Polaratom 1
end
```

The polarizability is calculated analytically, with the index "1" in the input block specifying this. Since DFT is the basis for electrostatics in the C-PCM approach used here, it serves as the reference method. Polarizabilities were calculated in a perfect conductor to keep the calculations consistent with the rest of the workflow. The resulting values are higher than those obtained in vacuum, but behave qualitatively the same.

The polarizability tensor is given by:

$$\alpha_{KL} = \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial r_K} \langle \mu | r_L | \nu \rangle. \tag{1}$$

Here, α_{KL} is the polarizability tensor ($K, L = x, y, z$), μ, ν are basis functions, $P_{\mu\nu}$ is the density matrix. The electric field E interacts with the dipole moment of the molecule as $-Ed$, where the dipole operator is given by:

$$d = \sum_A Z_A R_A - \sum_i r_i. \tag{2}$$

Here, A denotes nuclei at positions R_A with charge Z_A and i denotes electrons with position operator r_i. The derivative of the density matrix is the response density with respect to an electric field. In real space it reads:

$$\rho^{(K)}(\mathbf{r}) = \sum_{\mu\nu} \frac{\partial P_{\mu\nu}}{\partial r_K} \mu(\mathbf{r}) \nu(\mathbf{r}). \tag{3}$$

In order to arrive at a local polarizability for a given atom A, we decompose the real-space response density by multiplying it with a basic function $f^{(A)}(\mathbf{r})$ to get an atomic response density $\rho^{(A;K)}(\mathbf{r}) = f^{(A)}(\mathbf{r})\rho^{(K)}(\mathbf{r})$. The basic function must be chosen such that $\sum_A f^A = 1$. A particularly simple choice and the one we have adopted in this work is:

$$f^{(A)}(\mathbf{r}) = \frac{\tilde{\rho}^{(A)}(r)}{\sum_B \tilde{\rho}^{(B)}(r)}, \tag{4}$$

where, $\tilde{\rho}^{(A)}(r)$ is the spherically symmetric, tabulated density of a neutral atom A. Thus, in real-space the atomic polarizability tensor is defined as:

$$\alpha_{KL}^{(A)} = \int \rho^{(A;K)}(\mathbf{r}) r_L dr, \tag{5}$$

which fulfills the requirement that $\sum_A \alpha_{KL}^{(A)} = \alpha_{KL}$. The necessary integration is carried out numerically using the same grid as used in the preceding DFT calculation.

The "Polaratom" command in the input block decomposes the molecular polarizability tensor into atomic components.

Another method implemented with PSI4 for partitioning molecular polarizability tensors into atomic contributions is reported in Heid et al. (2018), where the partitioning is based on distributed multipole analysis of wavefunctions within each molecule. In addition to quantum

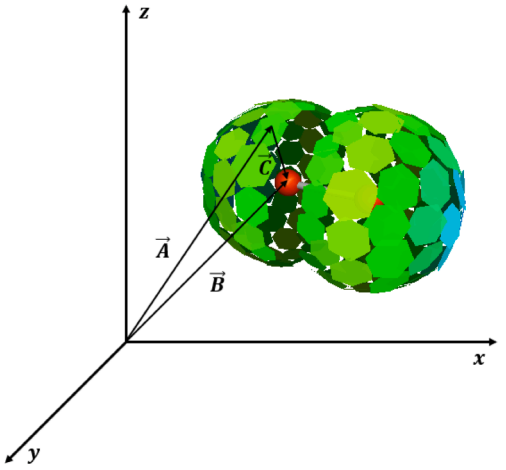

Fig. 1. The molecular cavity of a diatomic molecule (Br_2) with its segments and vectors connecting the center of the coordinate system and a segment center ($\vec{A}$), an atom center ($\vec{B}$), and the center of the segment to the atom center ($\vec{C}$). The molecular cavity for the image is created using COSMOtherm v.24.

chemical calculations, atomic polarizabilities can also be estimated using machine learning approaches, as demonstrated by Heid et al. (2019).

2.2. Atomic polarizability as a segment descriptor

For an initial assignment of polarizability values to each segment, we considered isotropic polarizabilities, α_{iso}, calculated from the diagonal elements ($\alpha_{xx}, \alpha_{yy}, \alpha_{zz}$) of an atomic polarizability tensor:

$$\alpha_{\text{iso}} = \frac{\alpha_{xx} + \alpha_{yy} + \alpha_{zz}}{3}. \tag{6}$$

Additionally, atomic polarizability tensors were projected onto segments. Using these as a new segment descriptor is inspired by the London attractive potential between two spherically symmetric systems (London, 1937):

$$\phi_{ij}^{\text{London}} = -\frac{3}{2}\frac{\alpha_i \alpha_j}{r_{ij}^6} \cdot \frac{h\nu_i h\nu_j}{h(\nu_i + \nu_j)}, \tag{7}$$

where α_i and α_j are their polarizabilities, and r_{ij} is the distance between them. Eq. (7) further demonstrates the effect of zero-point vibrations with frequencies ν_i and ν_j, which can be approximated by ionization potentials (Hudson and McCoubrey, 1960).

2.2.1. Projection of atomic polarizability tensors onto molecular cavities

Fig. 1 illustrates the projection of the atomic polarizability tensor onto the molecular cavity. Vector $\vec{A}$ connects the coordinate system center to the center of segment i, and vector $\vec{B}$ to the atom center. Vector $\vec{C}$, connecting these two centers, is calculated by subtracting the coordinates of $\vec{A}$ from those of $\vec{B}$. We first take the absolute value of the coordinates of $\vec{C}$ and then normalize $\vec{C}$ by the sum of its coordinates to equalize all vectors from other segments relative to their location with respect to the atom center. For each atom, ORCA electrical properties calculations provide a 3-dimensional polarizability tensor in the form of a 3×3 matrix. The projection of the atomic polarizability tensor $\boldsymbol{\alpha}$ onto the atom's segments is then performed via matrix multiplication of this tensor with vector $\vec{C}$:

$$\vec{D} = \vec{C} \cdot \boldsymbol{\alpha}. \tag{8}$$

To derive a scalar descriptor, we compute the Euclidean norm of the resulting vector $\vec{D}$, ensuring a physically meaningful positive value. This norm is then used as a local descriptor of the segment's polarizability, hereafter referred to as the polarizability projection α_i of segment i.

This procedure results in a large number of possible polarizability projection values within a single molecule, which significantly increases computational time due to the expanded variety of segment types. To keep computational time manageable, we performed the clustering of α_i by essentially rounding them up with further details in Section 3.1.3. In a similar manner as the screening surface charge densities are clustered to generate the σ-profiles in COSMO-RS, we can analogously plot α-profiles as illustrated in Fig. 2. It can be observed that as the atomic number of the halogen increases, the α-profiles of the corresponding haloalkanes shift toward higher polarizability projection values. This trend aligns with the expected physical properties of halogens and haloalkanes.

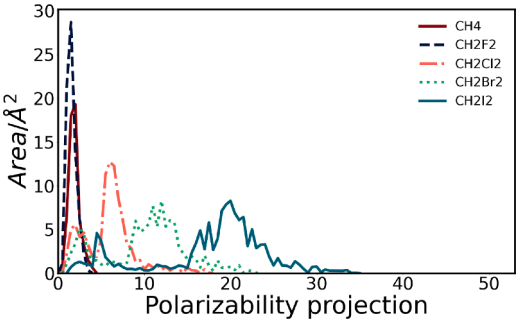

Fig. 2. α-profiles of methane and dihaloalkanes.

To incorporate more molecular information to the new descriptor, we evaluated five scaling factors or weightings ($w_1 - w_5$) for each projection α_i. Scaling factors correct the polarizability projections by incorporating other atomic properties to distinguish between the regions of the molecules with identical polarizabilities.

w_1: Normalization by the initial number of segments n_{seg}^I that atom I was fragmented into as a result of C-PCM calculations, expressed as

$$w_1 = \frac{1}{n_{\text{seg}}^I}. \tag{9}$$

w_2: The average number of segments across all atoms is used in Eq. (9).

w_3: The ratio of the segment area a_i to the total area of atom I:

$$w_3 = \frac{a_i}{A_I}. \tag{10}$$

w_4: The average atomic area across the molecule is used in the denominator of Eq. (10).

w_5: Based on the COSMO-RS model's σ-averaging method, this scaling substitutes σ with α_i:

$$\alpha_i^* = \frac{\sum_j^N \alpha_j \frac{r_j^2 r_{\text{av}}^2}{r_j^2 + r_{\text{av}}^2} \exp\left(\frac{-d_{ij}^2}{r_j^2 + r_{\text{av}}^2}\right)}{\sum_j^N \frac{r_j^2 r_{\text{av}}^2}{r_j^2 + r_{\text{av}}^2} \exp\left(\frac{-d_{ij}^2}{r_j^2 + r_{\text{av}}^2}\right)}. \tag{11}$$

In the first four cases, the scaled polarizability is obtained by simply multiplying α_i by one of the scaling factors, w_i. Using the scaling factors w_1 and w_3, we can distinguish between segments located on different atoms while sharing the same α_i. The scaling factors w_3 and w_4 also account for the area of an individual segment. Similar to the original σ-averaging procedure, Eq. (11) smears out segments and their corresponding α_i.

2.3. COSMO-RS

For a detailed explanation of the model, we refer the reader to foundational works by Klamt et al. (1998), Klamt and Eckert (2000) and to our previous studies (Gerlach et al., 2022; Grigorash et al., 2024). Here, we highlight only the aspects of the model that were modified or differ from the original implementation.

In COSMO-RS, thermodynamic properties are calculated based on the interaction energies between molecular segments $E_{ij}(i, j)$, with three contributing components:

$$E_{ij}(i,j) = E_{\mathrm{mf}}(\sigma_i,\sigma_j) + E_{\mathrm{hb}}(\sigma_i,\sigma_j) - E_{\mathrm{vdW}}(i,j). \quad (12)$$

These contributions include the repulsive misfit energy $E_{\mathrm{mf}}(\sigma_i,\sigma_j)$, the hydrogen-bonding energy $E_{\mathrm{hb}}(\sigma_i,\sigma_j)$, and the dispersive energy $E_{\mathrm{vdW}}(i,j)$, which is not a function of screening surface charge densities but is defined via other properties of interacting segments. Hydrogen bonding and dispersion are attractive interactions, therefore the corresponding energy terms either always take negative values in the case of $E_{\mathrm{hb}}(\sigma_i,\sigma_j)$, or are subtracted from the total segment-segment interaction energy, as with $E_{\mathrm{vdW}}(i,j)$. The repulsive misfit energy is typically calculated as:

$$E_{\mathrm{mf}}(\sigma_i,\sigma_j) = 0.5 a_{\mathrm{eff}} \alpha_{\mathrm{mf}} \left[(\sigma_i+\sigma_j)^2 + f_{\mathrm{corr}}(\sigma_i+\sigma_j)(\sigma_i^{\perp}+\sigma_j^{\perp}) \right], \quad (13)$$

which incorporates both the screening charge density σ and the correlation screening charge density $\sigma^{\perp}$ to account for the influence of the surrounding segments. Here, f_{corr} represents a correlation correction factor adjusted to dielectric energy data (Klamt et al., 1998). The general COSMO-RS parameters α_{mf} and a_{eff} are the misfit prefactor and the effective contact area of a segment, respectively.

In this study, to enhance predictions for fluorinated molecules, the misfit free energy was corrected by an additional repulsive energy term $E^{\mathrm{F}}_{\mathrm{corr}}$ to all segment-segment interactions involving an F-atom (F - segments). This modification leads to the following adjusted equation:

$$E_{\mathrm{mf}}(\sigma_i,\sigma_j) = 0.5 a_{\mathrm{eff}} \alpha_{\mathrm{mf}} \left[(\sigma_i+\sigma_j)^2 + f_{\mathrm{corr}}(\sigma_i+\sigma_j)(\sigma_i^{\perp}+\sigma_j^{\perp}) \right] + E^{\mathrm{F}}_{\mathrm{corr}}. \quad (14)$$

In the first part of this work, the dispersive energy contribution was incorporated as in Eq. (12). Each segment was assigned a parameter based on its atomic number, which was regressed using halocarbon thermodynamic data. A similar approach–using atomic number as a segment descriptor–is implemented in COSMOtherm (Klamt and Eckert, 2007) to improve the modeling of dispersive interactions. Another implementation, COSMO-SAC-dsp (Hsieh et al., 2014), also introduces atomic parameters to account for dispersion, though without explicitly linking them to segments. On the other hand, COSMO-SAC (Hsieh et al., 2010) treatment of hydrogen bonding involves differentiation of segments by atomic number to assign them to distinct hydrogen bond groups. This enables a more specific treatment of hydrogen bonding compared to earlier COSMO-RS implementations.

In the present study, the dispersive interactions between segments i and j are calculated based on a new descriptor, which is either the atomic isotropic polarizability of the corresponding atom or the polarizability projection:

$$E_{\mathrm{vdW}}(\alpha_i,\alpha_j) = a_{\mathrm{eff}} m_{\mathrm{vdW}} f(\alpha_i,\alpha_j). \quad (15)$$

For the polarizability combining function $f(\alpha_i,\alpha_j)$, we evaluated several forms:

$$f(\alpha_i,\alpha_j) = \sqrt{\alpha_i \times \alpha_j}, \quad (16)$$

$$f(\alpha_i,\alpha_j) = \alpha_i \times \alpha_j, \quad (17)$$

including a variant where the root power was treated as an adjustable parameter fitted to experimental data. Additionally, we considered a logarithmic combining function defined as:

$$f(\alpha_i,\alpha_j) = \ln\left(1 + \frac{\alpha_i \alpha_j}{\mathrm{a.u.}^6}\right). \quad (18)$$

Here, the constant "1" is added to $\alpha_i\alpha_j$ to ensure that the logarithm yields a zero value when one of the polarizability projections is zero. Additionally, $\alpha_i\alpha_j$ is divided by a.u.^{6} to maintain the dimensionless nature of the logarithmic argument, however, in the forthcoming discussion, we will omit it for simplicity. In certain cases, the dispersive interactions were differentiated between F - segments and all other segments. For this purpose, two distinct scaling factors were applied: $m^{\mathrm{F}}_{\mathrm{vdW}}$ for interactions involving an F-atom and m_{vdW} otherwise.

For the combinatorial contribution, we considered the Flory-Huggins (FH) term (Flory, 1942), following the assessment results of the first part of this study. Furthermore, it has been shown in Krooshof and de With (2024) that the FH term is the only physically consistent, purely combinatorial model among other commonly used combinatorial terms.

2.4. Regression procedure

To optimize the parameters of openCOSMO-RS, we considered various types of phase equilibrium data: VLE, LLE, solvation free energy (ΔG_{solv}), infinite dilution activity coefficients (IDAC), and mole fraction based partitioning coefficients of component i between aqueous and organic phases ($K_i^{\mathrm{org/aq}}$) estimated as:

$$K_i^{\mathrm{org/aq}} = \frac{\gamma_i^{\infty,\mathrm{aq}}}{\gamma_i^{\infty,\mathrm{org}}} \quad (19)$$

For VLE, LLE, IDAC and $K_i^{\mathrm{org/aq}}$, we followed the same data processing procedure and used the same objective function as in Part 1 of this work (Grigorash et al., 2024).

The solvation free energy, ΔG_{solv}, was calculated using:

$$\Delta G_{\mathrm{solv}} = E_{\mathrm{diel}} + RT\ln\gamma_i^{\infty} - \sum_I \tau_I A_I - \omega_{\mathrm{ring}} n_{\mathrm{ring}} - RT\ln\frac{\nu_{\mathrm{IG}}}{\nu_{\mathrm{liquid}}} - \eta, \quad (20)$$

where E_{diel} is the predicted dielectric energy associated with transferring the solute from the gas phase to an ideal conductor. It is calculated using ORCA 6.0 as the difference between the energy of the solute in a perfect conductor and in vacuum. The second term corresponds to the liquid phase chemical potential of the solute at infinite dilution, with the ideal conductor as the reference state, as predicted by openCOSMO-RS. The third term accounts for the energy associated with cavity formation, incorporating a van der Waals-like contribution calculated as the product of each atom's I surface area A_I and an atomic-number-dependent factor τ_I. To correct for ring-containing molecules, the fourth term introduces a contribution proportional to the number of rings n_{ring}, scaled by a general parameter ω_{ring}. The fifth term adjusts for the change in reference states from mole fractions to molar concentrations, with the molar volume of the ideal gas - ν_{IG} and of liquid phase - ν_{liquid}. Lastly, η is an adjustable parameter. The molar volumes were calculated using a predictive QSPR model developed by Müller et al. (2025). This model is based on molecular descriptors, such as σ–moments derived from σ–profiles, number of atoms and molecular surface area. Overall, the solvation free energy model involves 2 general ($\omega_{\mathrm{ring}}, \eta$) and 11 atomic parameters (τ_I).

The models' performance was assessed using the Average Absolute Deviation (AAD), calculated as:

$$\mathrm{AAD} = \frac{1}{N_p}\sum_{s=1}^{N_s} \overline{\left|\ln(X_i^{\mathrm{calc}}) - \ln(X_i^{\mathrm{exp}})\right|}, \quad (21)$$

where X_i represents $\ln\gamma_i^{\infty}$, $\ln\gamma_i^{\mathrm{VLE}}$, $K_i^{\mathrm{org/aq}}$, K_i^{LLE}, or ΔG_{solv} and N_p is the number of respective data points. The overline indicates an average for binary mixture properties when two measurements are available (e.g., $\ln\gamma_1^{\mathrm{VLE}}$, $\ln\gamma_2^{\mathrm{VLE}}$). The total $\mathrm{AAD}_{\mathrm{TOT}}$, was computed as the sum of absolute deviations across all data types, normalized by the total number of data points.

3. Results and discussion

In the following sections, we discuss the challenges associated with modeling intermolecular interactions in fluorocarbon mixtures and explore various approaches to address these issues. To guide the reader through the detailed development process, Fig. 3 presents a schematic overview of the methods used to incorporate atomic polarizabilities into openCOSMO-RS applied to halocarbon data. Initially, atomic isotropic polarizabilities are assigned to all segments of the respective atoms, and various combining rules (CR) are evaluated to determine the optimal number of adjustable parameters. To further refine this approach, polarizability tensors are projected onto molecular segments, allowing for

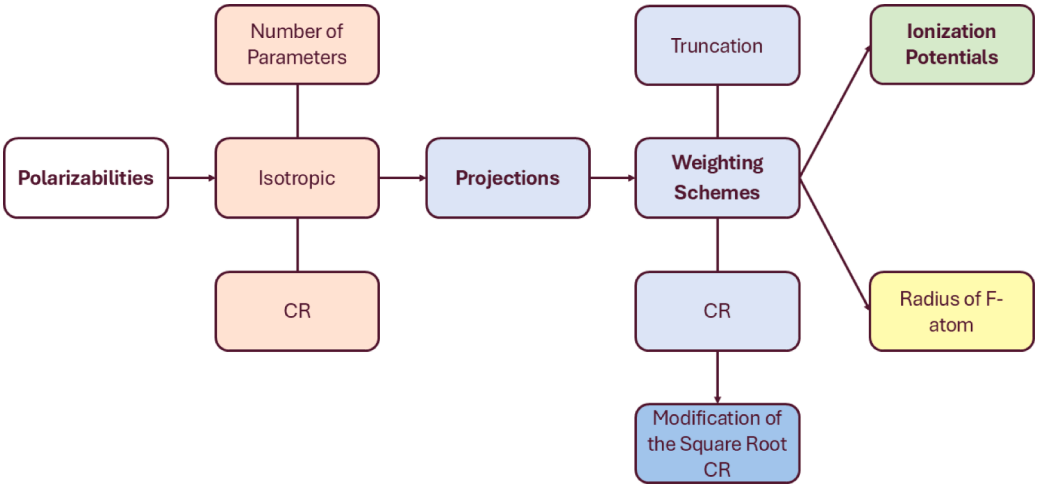

Fig. 3. Schematic workflow for incorporating atomic polarizabilities into openCOSMO-RS using halocarbon data.

a more localized estimation of polarizability across different regions of each atom. Multiple CR and correction strategies are then assessed to identify the most accurate and robust model. Following this, the most promising approaches (highlighted in bold in Fig. 3) were evaluated on the entire dataset, extending beyond halocarbons to ensure the general applicability of the model.

Table 1
An overview of the halocarbon collection of experimental data.

Type	Data points	Temperature range, K
VLE	10,104	183.01–463.25
LLE	1534	87.58–431.25
IDAC	2171	249.95–413.15

3.1. Halocarbons

3.1.1. Intermolecular interactions in fluorocarbon mixtures

The physical nature of intermolecular interactions between alkanes and fluoroalkanes is still not completely understood. Specifically, the thermodynamic behavior of these mixtures, such as their immiscibility, contradicts expectations based on their similar structure and polarity. Various explanations have been proposed (Scott, 1958; Rowlinson and Swinton, 1982; Morgado et al., 2017; Siebert and Knobler, 1971; Song et al., 2003; Murray et al., 2021), with a general consensus that the observed interactions are weaker than anticipated. For example, a molecular simulation study by Song et al. (2003), demonstrated that reducing intermolecular interactions involving fluorine by 25 % led to a significant improvement in the model's accuracy, indicating that these interactions are indeed weaker than expected. This finding aligns with an earlier review by Scott (1958), who attributed it to the highly repulsive nature of fluorine. Consequently, a standard square root CR alone is insufficient to adequately describe the intermolecular forces in fluorine-containing systems.

To enhance models such as COSMO-RS for fluorinated compounds, explicit consideration of dispersive interactions has been proposed, as these were not included in the original implementations. For instance, Mambo-Lomba and Paricaud (2021) improved the COSMO-SAC-dsp model by reducing the dispersion parameter for F-atom, which improved VLE modeling of refrigerant blends. In our prior work (Grigorash et al., 2024), we introduced atomic dispersion parameters, which substantially enhanced the accuracy of thermodynamic property predictions for fluorinated mixtures. A similar approach has been implemented in COSMOtherm, as noted in Klamt and Eckert (2007), Sachsenhauser et al. (2014).

To further examine interactions involving fluorinated species, we expanded our existing halocarbon dataset used in the first part of this study by incorporating 1861 additional IDAC data points from Gmehling et al. (2008) (Table 1).

To account for the weaker dispersion interactions associated with fluorinated species within openCOSMO-RS, we implemented two modifications detailed in Section 2.3. The first modification introduces repulsive energy (E^{F}_{corr}, [J/mol]) for segment-segment interactions involving F-atom in the misfit term, as shown in Eq. (14). This parameter was optimized using experimental data for fluorinated compounds. Alternatively, the second modification adjusts the dispersion term (Eq. (15)) by selectively reducing the strength of dispersive interactions involving F-atom relative to those with other atoms. This approach required regressing two dispersion parameters: m^{F}_{vdW} for interactions involving F-segments and m_{vdW} for all other interactions not accounting for $f(\alpha_i, \alpha_j)$ at this stage. As shown in Fig. 4 and detailed in the first section of Table 2, both modifications led to similar improvements for data groups containing fluorinated compounds. However, as may be expected, for non-fluorinated groups, there was no noticeable enhancement compared to the original model. Since the dispersion parameters for the respective atoms remained unchanged, any residual contributions to the activity coefficient calculations were effectively canceled out.

3.1.2. Atomic isotropic polarizabilities

To improve the modeling of all halocarbons, including both fluorinated and non-fluorinated species, we incorporated atomic isotropic polarizabilities into the dispersion term. Each segment i is assigned an isotropic polarizability α_i of a respective atom. We then applied the three previously mentioned combining rules, each with a single adjustable scaling factor, m_{vdW}, for all types of segment-segment interactions, irrespective of the atom to which they belong. It should be noted that the units of m_{vdW} depend on the chosen CR, as isotropic polarizabilities are expressed in atomic units cubed (a.u.^{3}). Hereafter, the units of m_{vdW} are as follows: for the multiplication CR, [J/Å^2/a.u.^{6}]; for the square-root CR, [J/Å^2/a.u.^{3}]; and for the logarithmic CR, [J/Å^2]. The results are summarized in Table 2. It can be seen that without any additional correction for F-atom interactions, the overall improvement across all data categories remains modest compared to models incorporating the E^{F}_{corr} parameter, regardless of the combining rule applied. However, models that integrate both the E^{F}_{corr} correction in the misfit term and isotropic polarizabilities in the dispersion term show consistent improvements across all combining rules, with an example for the square root CR depicted in Fig. 5. There is a consistent improvement of AADs in groups containing Cl-, Br-, or I- atoms. For data groups involving F-atoms, however, the results are more controversial, which may indicate a deviation from the square root CR. For instance, the AAD for the VLE data

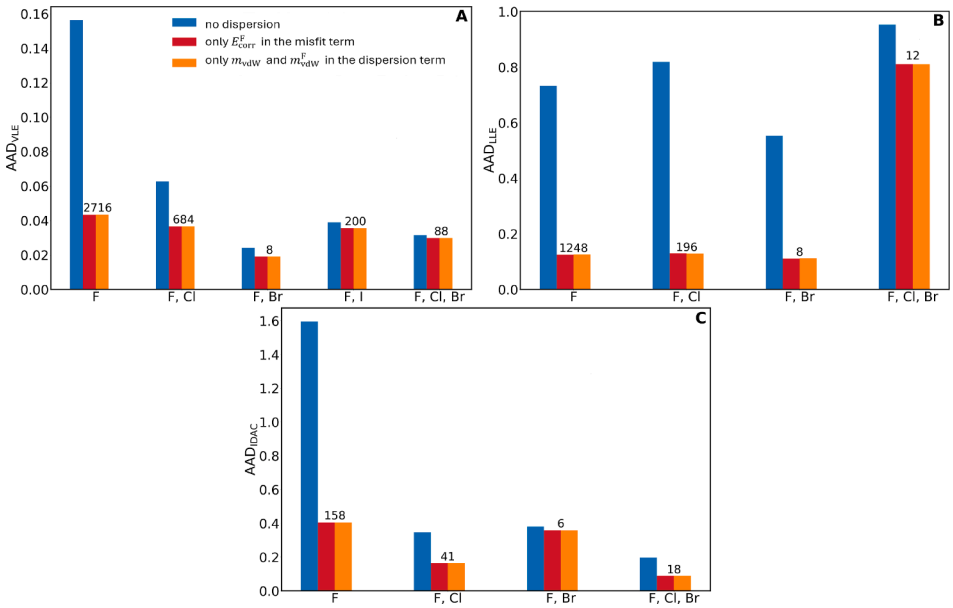

Fig. 4. Bar plots comparing the AAD for VLE (**A**), LLE (**B**) and IDAC (**C**) data across three models: the original model openCOSMO-RS 24a, and models incorporating corrections for F-atom interactions either through the misfit term or the dispersion term. The evaluated data is categorized based on the types of halogen atoms present in the molecules. Numbers of data points are indicated above columns for each category.

Table 2
Overview of various models, showing AADs and parameter values. Atomic isotropic polarizabilities are integrated into the dispersion term.

Model[a]			Parameters[b]			AAD			
the misfit correction	the dispersion term	CR	E^F_{corr}	m_{vdW}	m^F_{vdW}	TOT	IDAC	LLE	VLE
no dispersion/the original model[c]						0.194	0.415	0.727	0.068
✓			327.76			0.091	0.320	0.142	0.035
	✓			388.494	319.457	0.091	0.320	0.142	0.035
	✓			3.271		0.179	0.361	0.690	0.063
✓	✓		327.76	2.729		0.079	0.266	0.133	0.032
✓	✓	Eq. (16)	321.24	2.738		0.078	0.266	0.125	0.032
✓	✓		291.92	2.709	1.813	0.078	0.265	0.126	0.032
	✓			5.026	2.640E−06	0.132	0.348	0.389	0.048
	✓			0.042		0.179	0.350	0.704	0.064
✓	✓		327.76	0.037		0.078	0.259	0.132	0.032
✓	✓	Eq. (17)	323.50	0.037		0.077	0.260	0.127	0.032
✓	✓		313.78	0.037	2.0E−04	0.077	0.259	0.128	0.032
	✓			0.050	1.066E−07	0.174	0.350	0.668	0.063
	✓			779.954		0.179	0.388	0.692	0.059
✓	✓		327.76	444.426		0.080	0.270	0.148	0.031
✓	✓	Eq. (18)	316.85	458.098		0.079	0.271	0.138	0.030
✓	✓		181.75	459.129	451.599	0.077	0.269	0.129	0.030
	✓			472.255	454.964	0.078	0.273	0.138	0.029

[a] Model characteristics involve: 1) the misfit correction for F-segments using E^F_{corr} parameter (Eq. (14)); 2) the dispersion term (Eq. (15)), distinguishing F-interactions is indicated by non-zero value of m^F_{vdW}; 3) the combining rule for polarizabilities used in a dispersion term.

[b] The general parameters for openCOSMO-RS 24a with FH combinatorial term ($a_{eff} = 4.754$ [Å^2], $\alpha_{mf} = 7287$ [kJ/(mol · Å^2)/e^2], $c_{hb} = 43345$ [kJ/(mol · Å^2)/e^2], $\sigma_{hb} = 0.008918$ [e/Å^2], $r_{av} = 0.5$ [Å], $f_{corr} = 2.4$, $c^T_{hb} = 1.5$) were regressed using the IDAC, partitioning coefficients and solvation free energies from Müller et al. (2025). These general parameters are used for all evaluations of halocarbons. New parameters from Eqs. (14), (15).

[c] The original openCOSMO-RS 24a model.

involving both F- and I-atoms increases with the inclusion of polarizabilities, while it decreases for the F- and Br-atom group across most data types. Other groups, such as those with F and Cl, Cl and I, or F, Cl, and Br, show improvement in two out of three data types but worsen in the third. Nevertheless, the addition of the dispersion term with isotropic polarizabilities improves the overall AADs, as well as those for specific data types.

Furthermore, in this approach, both E^F_{corr} and m_{vdW} can be regressed simultaneously, which leads to a slight improvement in the overall results, as shown in third rows of Table 2 for each CR. However, this also increases the regression time. Therefore, in the subsequent model evaluations, we fix the E^F_{corr} parameter, and for the best-performing models, all other parameters will be adjusted simultaneously later. Additionally, segment-segment interactions involving F-atoms can be distinguished in the dispersion term by using a distinct scaling factor m^F_{vdW}. Subsequently, all three parameters–E^F_{corr}, m_{vdW}, and m^F_{vdW}–were adjusted simultaneously. Although this approach demonstrates superior AADs, the improvement over adjusting only two parameters is minimal. Therefore, the third parameter will not be used in further evaluations.

As we showed earlier, the additional repulsive energy in the misfit term can be replaced with two adjustable parameters in the dispersion term, leading to equivalent results. However, this is not feasible using either the square root or multiplication CRs. As shown in the last rows

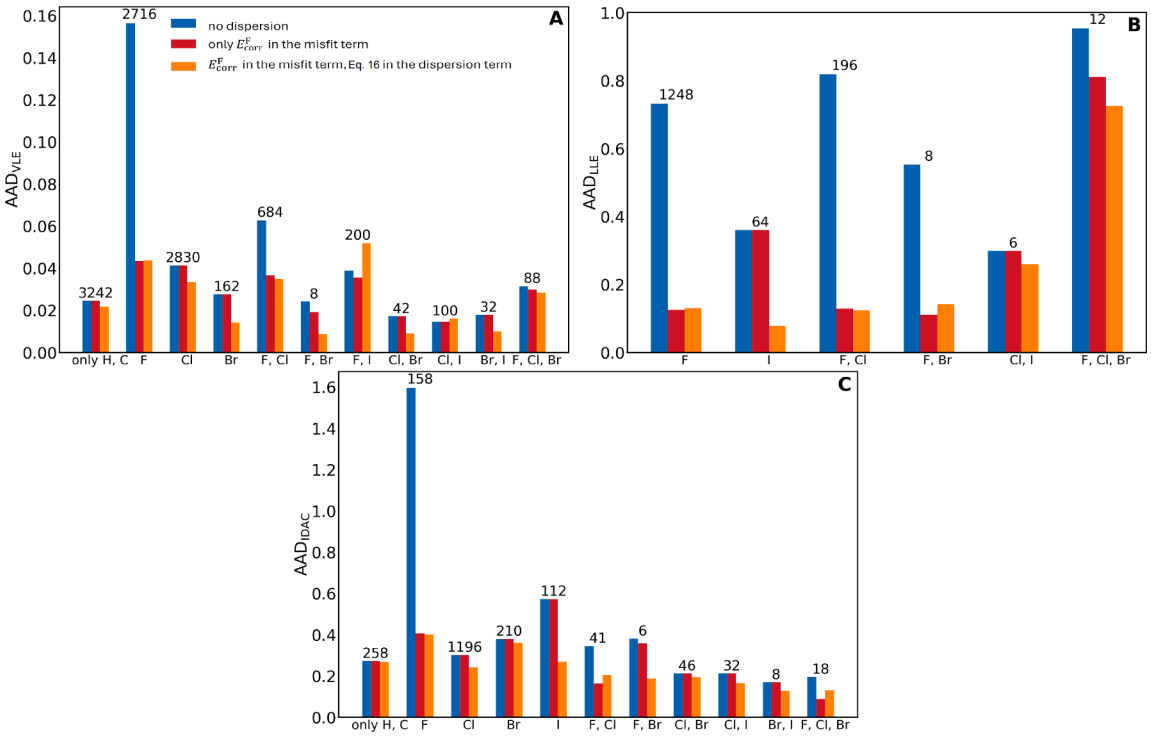

Fig. 5. Bar plots comparing the AAD for VLE (**A**), LLE (**B**) and IDAC (**C**) data across three models: the original model openCOSMO-RS 24a, the model incorporating corrections for F-atom interactions through the misfit term alone, and the model including both the misfit correction and a dispersion term with isotropic atomic polarizabilities. The evaluated data is categorized based on the types of halogen atoms present in the molecules. Numbers of data points are indicated above columns for each category.

corresponding to these CR in Table 2, the scaling factor for interactions involving the F-atom approaches zero, leading to higher total AADs compared to those achieved with the misfit term correction. On the other hand, this approach can be applied with the logarithmic combining rule, which smoothens the difference between α_i and α_j to the greatest extent among the considered combining rules. To understand the effect of the CR for the polarizabilities on the dispersion energy contribution to the interaction matrix, we evaluated the 3D dispersion energy surfaces for the investigated molecules. For instance, Fig. 6 illustrates such a surface for the $C_2HBrClF_3$ molecule, which includes nearly all halogens and thus covers a broad range of atomic polarizabilities. The surfaces were calculated for all possible pairs of α_i and α_j, using the scaling factors m_{vdW}, regressed with a fixed E^F_{corr} from Table 2. Noticeably, the energy range across these surfaces varies substantially, with the multiplication and square root CR resulting in lower energy values compared to the logarithmic combining rule. Nevertheless, this variation is not critical, as the activity coefficient represents a residual property. Therefore, the difference in the absolute values of $E_{vdW}(\alpha_i, \alpha_j)$ will be negated when subtracting the reference state activity coefficient.

3.1.3. *Local polarizability projections*

Another approach, described in Section 2.2.1, involves projecting the atomic polarizability tensors onto the molecular surface. This method assigns a local polarizability to each segment, as opposed to using a single isotropic atomic polarizability for individual atoms. In addition, we assessed different scaling strategies for the local polarizabilities of the segments that may account for the deviations from the chosen combining rule.

In analogy to clustering of the screening surface charge densities, we determined the optimal grid or the bin size for polarizability projections by performing regressions with the fixed $E^F_{corr} = 327.76$ and adjusting the scaling factors m_{vdW} for each α_i grid. The magnitude of the α_i values depends on the chosen scaling factor. To explore this, we considered two groups of grids. When α_i is unscaled or scaled using Eq. (11) (with α_i not exceeding 160), we tested grid step sizes of 1, 2, 5, 10, and 20. For scaling factors 1–4, which reduce α_i by approximately two orders of magnitude, we used step sizes of 0.01, 0.02, 0.05, 0.1, 0.2, and 0.5. This evaluation was performed for all CRs described earlier. Fig. 7 illustrates the resulting total AADs across all data types for different grid steps.

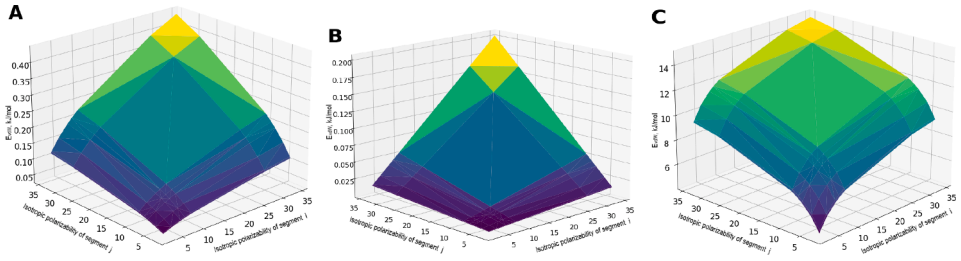

Fig. 6. The 3D dispersion energy $E_{vdW}(\alpha_i, \alpha_j)$ surfaces as functions of atomic isotropic polarizabilities for the $C_2HBrClF_3$ molecule, calculated using the square root (**A**), multiplication (**B**), and logarithmic (**C**) combining rules.

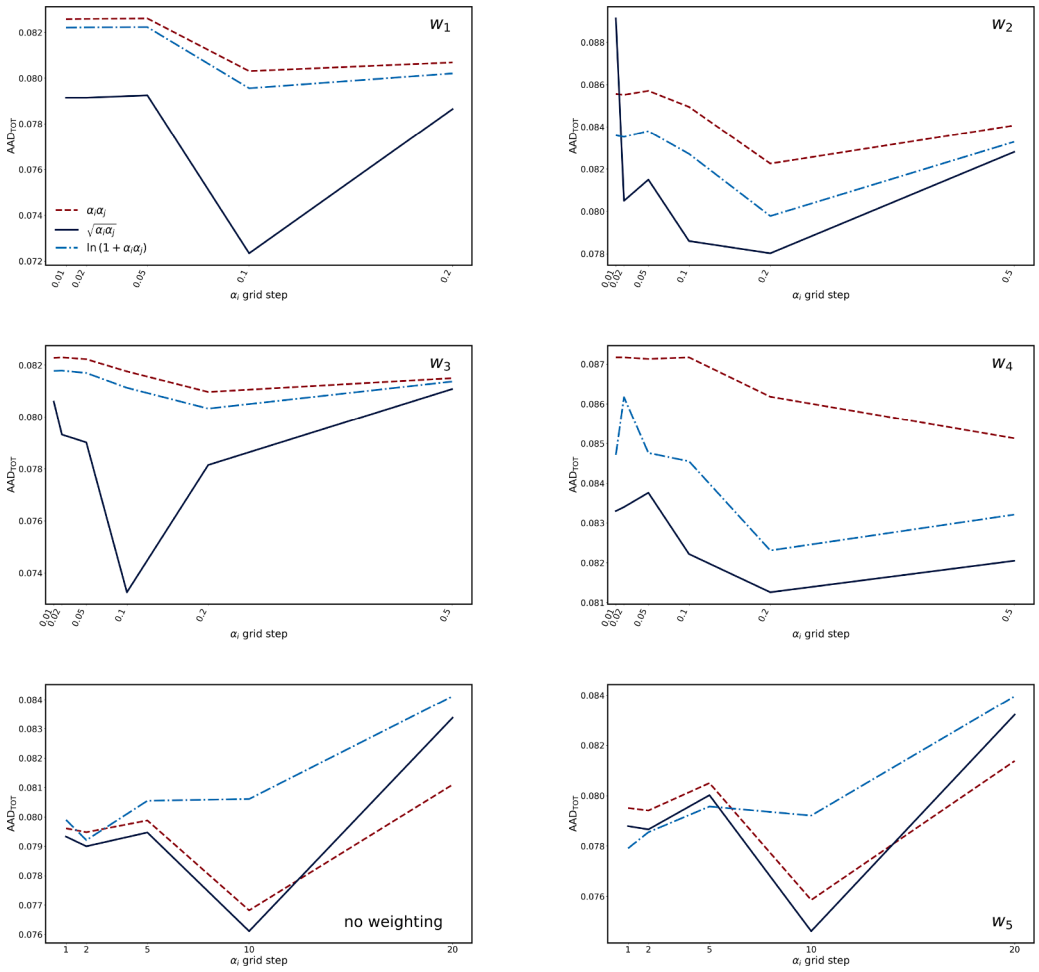

Fig. 7. The total AADs calculated for models, involving different scaling factors ($w_1 - w_5$) and CRs as a function of α_i grid step sizes.

Surprisingly, the finest grids do not always provide the most accurate predictions. Instead, optimal grids are often those that are 10 to 20 times coarser. To examine the factors behind this trend, we plotted the probability $p(\alpha_i)$ of finding a segment with a specific unscaled α_i on atoms (H, C, F, Cl, Br, or I) across all halocarbon molecules. In Fig. 8A, the segments were clustered using a grid step of 1, and in Fig. 8B, a grid step of 10 was used. Both plots show the expected polarizability trends for the atoms, however, with the coarser grid, nearly half of the segments associated with H-, C-, and F-atoms yield α_i values of zero. This suggests that the current version of openCOSMO-RS might implicitly account for the dispersion term within other terms, as a result of adjusting general model parameters without an explicit dispersion term. This could lead to double-counting of dispersion interactions, particularly for the most abundant atoms in the halocarbon systems (H, C, and F). This issue can be further addressed by simultaneously tuning the misfit parameters and dispersion parameters across a larger data set. A coarser grid also substantially reduces computational time by decreasing the number of segment types per system. Consequently, in further evaluations, we will focus on the coarse grids that yielded the lowest overall AADs. Fig. 7 indicates that using the square root CR along with E^{F}_{corr} in the misfit term might be the optimal modeling approach at this stage. In addition, the AADs across various scalings of α_i do not differ drastically. However, scaling factors 2 and 4 lead to higher deviations compared to 1 and 3, regardless of the CR or grid step size. This shows that the scaling factor should likely depend on the quantity of a particular atom, rather than the average value across the entire molecule (Eqs. (9) and (10)). Therefore, scaling factors 2 and 4 will not be considered within the upcoming analysis.

Table 3
Comparison of modeling approaches incorporating local polarizability projections with different scalings (w_i) on coarse α_i grids. The table includes AADs (Total, IDAC, LLE, VLE) and corresponding parameter values (E^{F}_{corr}, m_{vdW} from Eqs. (14), (15)).

Model		Parameters		AAD			
CR	w_i	E^{F}_{corr}	m_{vdW}	TOT	IDAC	LLE	VLE
Eq. (16)	None	327.76	3.139	0.076	0.257	0.139	0.029
	1 (Eq. (9))		811.338	0.072	0.239	0.124	0.030
	3 (Eq. (10))		716.554	0.073	0.250	0.127	0.029
	5 (Eq. (11))		3.090	0.075	0.242	0.145	0.029

The model parameters and AADs for the best-performing approaches, identified based on the total AADs shown in Fig. 7, are provided in Table 3. These models, which incorporate the square root CR and use coarse grids with step sizes of 0.1 for scaling factors 1 and 3, and 10 for the others, form the basis for further assessments. Compared to the approach with only E^{F}_{corr} in the misfit term, the total AADs of all models in Table 3 are reduced by 16–20 %, along with consistent improvements in AADs for specific data types. Furthermore, this approach provides better results than the one based on isotropic polarizabilities. This may result from considering local properties, using an effective scaling strategy, and selecting an appropriate grid step size.

3.1.4. *The evaluation of the F-atom radius used in C-PCM calculations*

Currently, the radius of the F-atom used for C-PCM calculations is set to the same value as that of the O-atom, as originally proposed by

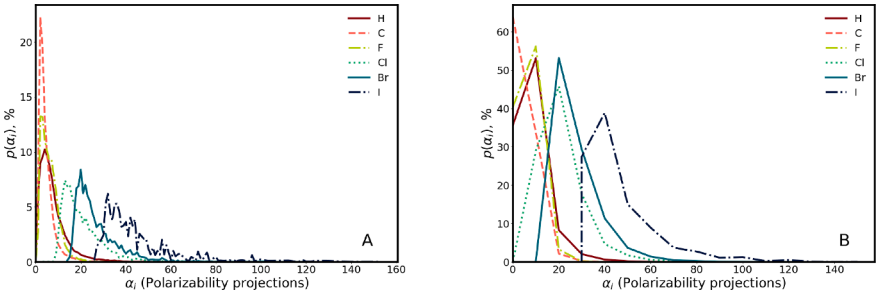

Fig. 8. The distribution of unscaled α_i ($p(\alpha_i)$) across all segments assigned to a particular atom (H, C, F, Cl, Br or I) in halocarbon molecules within the collected data. The clustering was performed using α_i grid step of 1 (**A**) and of 10 (**B**).

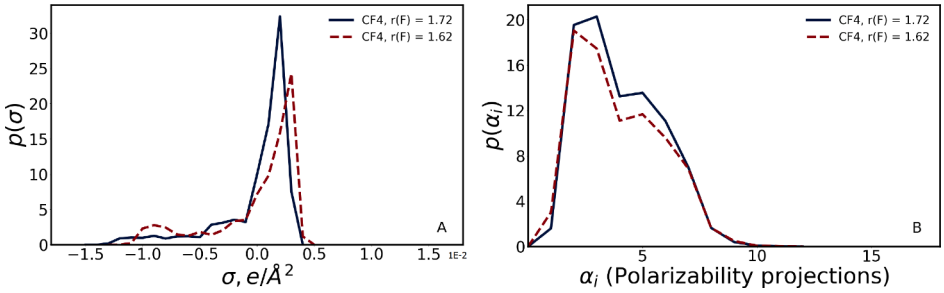

Fig. 9. Comparison of the $\sigma-$ profile (**A**) and $\alpha-$profile (**B**) of CF_4, calculated with both the original and reduced radii of the F-atom.

Klamt and Eckert (2000). However, from a physical perspective, the radius of the F-atom should be smaller. To investigate whether this discrepancy could be contributing to modeling issues in fluorinated mixtures, we performed the same C-PCM calculations for fluorinated molecules, but with the radius of the F-atom reduced from 1.72 to 1.62 Å. The same reduction of 6 % is observed in the C-PCM radius of the O-atom relative to the N-atom. As shown in the σ-profile of CF_4 illustrated in Fig. 9A, the reduced radius of the F-atom leads to a slight increase in the molecule's polarity, and as expected, to a decrease in its total area. On the other hand, the α-profile in Fig. 9B do not indicate any significant change in the distribution of polarizability.

For the updated C-PCM files, tuning the $E^{\mathrm{F}}_{\mathrm{corr}}$ parameter resulted in its increase from 327.76 to 354.625, compared to the results with the old radius. Additionally, the total AAD decreased by approximately 0.5 %. Subsequently, the scaling factors for the previously selected scaling approaches were regressed along with the updated $E^{\mathrm{F}}_{\mathrm{corr}}$ parameter (Table 4). Although there is some improvement for unscaled polarizability projections and those scaled with w_5 (Eq. (11)), the other two models show slightly inferior results. Generally speaking, despite the physically justified reduction in the F-atom radius, no clear advantage is observed over Klamt's conventional value.

3.1.5. Modification of the square root combining rule for polarizability projections

Dispersive interactions do not always align perfectly with the square root CR, therefore, alternative approaches have been proposed to address this limitation. One such approach, suggested in the molecular simulation study by Gould and Bučko (2016) of dispersive interactions between individual atoms, involves optimizing the power of the root used to combine the dispersion coefficients C_6. We have adopted this technique for polarizability projections, using CR with an adjustable root power β:

$$f(\alpha_i, \alpha_j) = (\alpha_i \times \alpha_j)^{\beta}, \tag{22}$$

for α_i, we selected scaling factors that showed the best performance in previous evaluations.

Interestingly, the parameter β regressed using unscaled α_i, shown in Table 5, is very similar to the power of 0.73 obtained in Gould and Bučko (2016). Although this approach appears somewhat empirical, it may confirm the fact that combining quantities related to the dispersion energy requires more complexity than the conventional square root CR. As demonstrated in Table 5, a minor improvement is observed across all scalings, except for w_1. This is likely because both the scaling and the adjusted power of the root aim to account for the deviations from the square root CR, and using both could potentially lead to error cancellation or accumulation.

3.1.6. Further refinement based on the atomic ionization potentials

Dispersive interactions between two spherical atoms are influenced by their ionization potentials (IPs) (London, 1937; Prausnitz et al., 1999), which quantify the energy required to remove an electron from an atom. This relationship has been employed in the development of

Table 4
Overview of modeling approaches based on C-PCM calculations with a reduced F-atom radius of 1.62 Å. The table includes AADs (Total, IDAC, LLE, VLE) and corresponding parameter values ($E^{\mathrm{F}}_{\mathrm{corr}}$, m_{vdW}, and $m^{\mathrm{F}}_{\mathrm{vdW}}$ from Eqs. (14), (15)). Local polarizability projections with different scalings (w_i) on coarse α_i grids are used in the dispersion term.

Model	w_i	Parameters			AAD			
		$E^{\mathrm{F}}_{\mathrm{corr}}$	m_{vdW}	$m^{\mathrm{F}}_{\mathrm{vdW}}$	TOT	IDAC	LLE	VLE
no dispersion term		354.625			0.090	0.319	0.144	0.034
no misfit correction			317.394	242.785	0.090	0.319	0.145	0.034
Eq. (16)	None	354.625	3.165		0.075	0.255	0.140	0.029
	1 (Eq. (9))		816.664		0.073	0.238	0.129	0.031
	3 (Eq. (10))		753.313		0.074	0.247	0.133	0.029
	5 (Eq. (11))		3.116		0.074	0.241	0.144	0.029

Table 5
Comparison of modeling approaches with adjusted β from Eq. (22), including their AADs (Total, IDAC, LLE, VLE) and corresponding parameter values (E^{F}_{corr}, m_{vdW}, β from Eqs. (14), (15), (22)). Local polarizability projections with various scalings (w_i) on coarse α_i grids are used in the dispersion term.

Model		Parameters			AAD			
CR	w_i	E^{F}_{corr}	m_{vdW}	β	TOT	IDAC	LLE	VLE
Eq. (22)	None	327.76	0.845	0.6918	0.0755	0.255	0.136	0.0294
	1 (Eq. (9))		415.881	0.2594	0.0757	0.232	0.130	0.0353
	3 (Eq. (10))		339.335	0.1418	0.0729	0.237	0.132	0.0301
	5 (Eq. (11))		1.928	0.5707	0.0741	0.242	0.143	0.0289

Table 6
Comparison of modeling approaches incorporating atomic ionization potentials into the dispersion term. The table includes AADs (Total, IDAC, LLE, VLE) and corresponding parameter values (E^{F}_{corr}, m_{vdW} from Eqs. (14), (15)) for local polarizability projections using different scalings (w_i) on coarse α_i grids.

Model		Parameters		AAD			
CR	w_i	E^{F}_{corr}	m_{vdW}	TOT	IDAC	LLE	VLE
Eq. (23)	None	327.76	3.126	0.076	0.256	0.140	0.029
	5 (Eq. (11))		3.082	0.075	0.242	0.146	0.029
Eq. (24)	None	327.76	38.652	0.074	0.248	0.130	0.029
	5 (Eq. (11))		37.709	0.072	0.234	0.135	0.029
Eq. (25)	None	327.76	77.091	0.074	0.250	0.131	0.029
	5 (Eq. (11))		75.135	0.072	0.234	0.136	0.029

combining rules for intermolecular interactions. For instance, IPs—often significantly different between species such as hydrocarbons and fluorocarbons–have been incorporated into the CRs proposed by Hudson and McCoubrey (1960). This approach has also been extended to predictive frameworks, such as the Group-Contribution Statistical Associating Fluid Theory model, where adjusted pseudo-ionization potentials of molecular groups are used (Nguyen-Huynh et al., 2008a,b).

Incorporating IPs into the dispersion term of openCOSMO-RS could enhance the accuracy of predictions, particularly when the IPs of interacting species differ significantly. For halocarbons, the most pronounced variation is between the IP of the F-atom (17.423) and the IPs of other atoms, which range from 10 to 13.5 as reported in (National Center for Biotechnology Information, 2024) for isolated atoms and directly used in our calculations. Although this difference is not exceedingly large, we evaluated the following equations incorporating atomic IPs (I_I, I_J) into segment-segment dispersion energy:

$$E_{vdW}(\alpha_i,\alpha_j) = a_{eff} m_{vdW} \sqrt{\alpha_i \alpha_j}\, \frac{2\sqrt{I_I I_J}}{I_I + I_J}; \tag{23}$$

$$E_{vdW}(\alpha_i,\alpha_j) = a_{eff} m_{vdW} \sqrt{\frac{\alpha_i \alpha_j}{I_I I_J}}; \tag{24}$$

$$E_{vdW}(\alpha_i,\alpha_j) = a_{eff} m_{vdW} \frac{\sqrt{\alpha_i \alpha_j}}{I_I + I_J}. \tag{25}$$

Only unscaled α_i values or those scaled with w_5 are reported, since other scalings yielded inferior AADs, potentially due to physically inconsistent overlaps of multiple corrections to the square root CR. For example, as atomic radius increases, ionization potential typically decreases. While the atomic radius influences scaling factors w_1(Eq. (9)) and w_3 (Eq. (10)), this effect opposes that of the ionization potential on α_i. Consequently, applying both corrections simultaneously can lead to partial or full cancellation of their effects, reducing the physical consistency and predictive accuracy of the model.

Eq. (23) is based on the predictive approach for estimating the k_{ij} parameter proposed in (Hudson and McCoubrey, 1960), but its effect on the AADs is minimal, as the correction factor tends to one. However, this method might be more applicable to systems with greater differences in IPs. Additionally, we did not account for local IPs, which differ more profoundly within a molecule (Jin et al., 2004) and therefore may be more sensitive to this approach. In Eq. (24), IPs serve as scalings to α_i, while Eq. (25) considers only the denominator of Hudson's k_{ij}. These two equations produced the best results when w_5 was applied to α_i. While the mathematical formulation remains somewhat empirical–much like certain aspects of the model itself–it retains a physical basis. Hence, we observe the improvement in model's predictions without introducing additional adjustable parameters, but solely by integrating atomic IPs.

3.1.7. *Concluding remarks on incorporating polarizabilities for modeling halocarbon thermodynamic properties*

In the preceding discussion, we evaluated several approaches to incorporate polarizability as a first-principles molecular descriptor. Among the various combining rules, the square root emerged as the optimal choice over conventional alternatives. However, it requires corrections such as scalings, ionization potentials, or adjusting the root power. Notably, combining multiple corrections can sometimes yield inferior results. Based on prior assessments, the scaling factor w_1 and the scaling factor w_5, coupled with IPs using Eqs. (24) or (25), demonstrated the best performance.

To further refine the methodology, we simultaneously regressed two key parameters: the misfit correction E^{F}_{corr} for interactions involving the F-atom, and the m_{vdW} parameter, scaling the dispersion term. Additionally, we compared this new model with the FH_6 approach developed in the first part of this study. To ensure a fair comparison, we re-optimized the dispersion atomic parameters using the updated dataset and the ORCA 6.0 C-PCM files. As in previous analyses, optimization involved the differential evolution algorithm (Virtanen et al., 2020), with physically reasonable boundaries. Table 7 summarizes the comparison. The new approach, despite containing three times fewer adjustable parameters, demonstrated superior performance. We further assessed the models' accuracy across different data groups. Fig. 10 reveals that the FH_6 model underperforms for fluorinated groups, highlighting the importance of the misfit correction in addition to the dispersion term. For other non-fluorinated, halogenated groups, the dispersion term incorporating α_i generally performs similarly to one with individual

Table 7
Comparison of the new modeling approaches, incorporating α_i and simultaneously adjusted E^{F}_{corr} and m_{vdW} from Eqs. (14), (15) with the model based on atomic dispersion parameters (FH_6). The table includes AADs (Total, IDAC, LLE, VLE) and corresponding parameter values.

Model		Parameters		AAD			
CR	w_i	E^{F}_{corr}	m_{vdW}	TOT	IDAC	LLE	VLE
Eq. (16)	1 (Eq. (9))	304.92	842.20	0.069	0.238	0.102	0.029
Eq. (16)	5 (Eq. (11))	310.64	3.141	0.072	0.242	0.125	0.029
Eq. (24)	5 (Eq. (11))	313.85	38.044	0.070	0.234	0.118	0.029
Eq. (25)	5 (Eq. (11))	312.62	76.165	0.070	0.234	0.118	0.029
FH_6[a]				0.077	0.250	0.133	0.033

[a] Atomic dispersion parameters [$J^{0.5}/Å$]: $\tau^{vdW}_{H} = 6.745$; $\tau^{vdW}_{C} = 14.719$; $\tau^{vdW}_{F} = 2.181$; $\tau^{vdW}_{Cl} = 12.654$; $\tau^{vdW}_{Br} = 16.167$; $\tau^{vdW}_{I} = 21.211$.

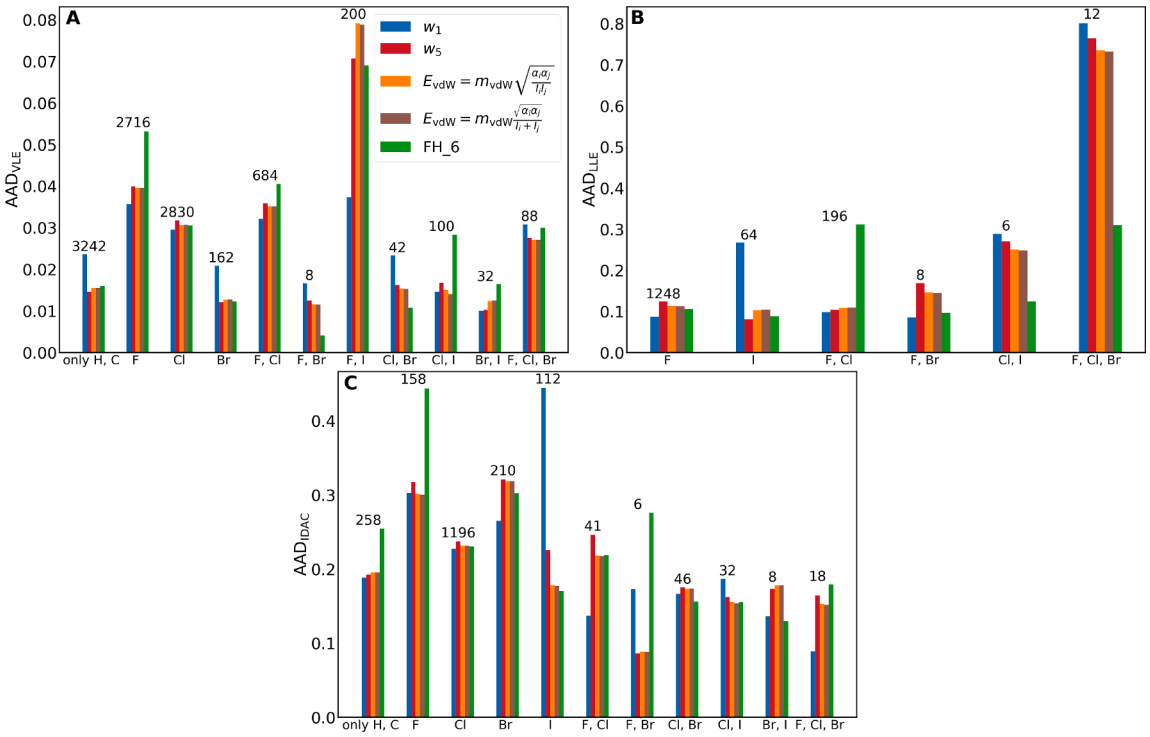

Fig. 10. Bar plots comparing the AAD for VLE (**A**), LLE (**B**) and IDAC (**C**) data across five models: the models incorporating the dispersion term with different scalings and corrections for α_i along with the E^{F}_{corr} in the misfit term, and the model based on atomic dispersion parameters (FH_6). The evaluated data is categorized based on the types of halogen atoms present in the molecules. Numbers of data points are indicated above columns for each category.

Table 8
An overview of the complete collection of experimental data.

Type	Data points	Temperature range, K	Sources
VLE	37,730	183.01–573.15	Grigorash et al. (2024), KDB (2024)
LLE	6882	87.58–593	Grigorash et al. (2024), Riccardi et al. (2021)
ΔG_{solv}	2126	298.15	Müller et al. (2025)
IDAC	6763	248.15–473.05	Gerlach et al. (2022), Gmehling et al. (2008)
$K_i^{org/aq}$	296	298.15	Gerlach et al. (2022)

Table 9
Comparison of the new modeling approaches, incorporating α_i with simultaneously adjusted E^{F}_{corr}, m_{vdW} from Eqs. (14), (15), and model's general parameters. The table includes AADs (Total, $K_i^{org/aq}$, IDAC, LLE, VLE, ΔG_{solv}) and corresponding parameter values. Parameters that were fixed during optimization are marked with asterisk.

Parameters and Deviations (AAD)	Model				
	no dispersion	Eq. (16) + w_1 (Eq. (9))	Eq. (16) + w_5 (Eq. (11))	Eq. (24) + w_5 (Eq. (11))	Eq. (25) + w_5 (Eq. (11))
a_{eff} [Å^2]	4.574	4.943	4.955	4.90825	4.706
r^*_{av} [Å]			0.5		
α_{mf} [kJ/(mol·Å^2)/e^2]	7613	7294	7672	7876	7322
f^*_{corr}			2.4		
c_{hb} [kJ/(mol·Å^2)/e^2]	53,447	43,265	43,457	49,318	43,421
c^{T*}_{hb}			1.5		
σ_{hb} [e/Å^2]	0.009739	0.009492	0.00960	0.009953	0.009355
E^{F}_{corr} [J/mol]		349.18	350.37	346.82	340.73
m_{vdW}		339.58	2.4097	29.567	27.853
AAD_{TOT}	0.216	0.178	0.172	0.169	0.1795
$AAD_{K_i^{org/aq}}$	0.858	0.790	0.806	0.812	0.831
AAD_{IDAC}	0.570	0.464	0.461	0.452	0.476
AAD_{LLE}	0.396	0.258	0.248	0.244	0.265
AAD_{VLE}	0.102	0.093	0.086	0.084	0.091
$AAD_{\Delta G_{solv}}$	0.444	0.435	0.444	0.443	0.443

Table 10
Extension of Table 9 with parameters for calculating ΔG_{solv} (Eq. (20)).

Parameters	Model				
	no dispersion	Eq. (16) + w_1 (Eq. (9))	Eq. (16) + w_5 (Eq. (11))	Eq. (24) + w_5 (Eq. (11))	Eq. (25) + w_5 (Eq. (11))
τ_1 [$\frac{kJ}{mol \cdot Å^2}$]	0.027	0.028	0.031	0.030	0.029
τ_6 [$\frac{kJ}{mol \cdot Å^2}$]	0.025	0.016	0.019	0.019	0.021
τ_7 [$\frac{kJ}{mol \cdot Å^2}$]	0.013	0.003	0.006	0.009	0.007
τ_8 [$\frac{kJ}{mol \cdot Å^2}$]	0.008	0.003	0.006	0.010	0.005
τ_9 [$\frac{kJ}{mol \cdot Å^2}$]	0.007	0.018	0.025	0.019	0.022
τ_{17} [$\frac{kJ}{mol \cdot Å^2}$]	0.033	0.028	0.028	0.029	0.031
τ_{35} [$\frac{kJ}{mol \cdot Å^2}$]	0.041	0.035	0.035	0.035	0.039
τ_{53} [$\frac{kJ}{mol \cdot Å^2}$]	0.731	0.531	0.321	1.530	0.384
τ_{14} [$\frac{kJ}{mol \cdot Å^2}$]	0.057	0.016	0.038	0.005	0.016
τ_{15} [$\frac{kJ}{mol \cdot Å^2}$]	0.020	0.002	0.004	0.0001	0.007
τ_{16} [$\frac{kJ}{mol \cdot Å^2}$]	0.035	0.027	0.026	0.026	0.031
η [$\frac{kJ}{mol}$]	−4.271	−4.086	−4.134	−4.102	−4.252
ω_{ring} [$\frac{kJ}{mol}$]	0.2423	0.2351	0.2511	0.2495	0.2491

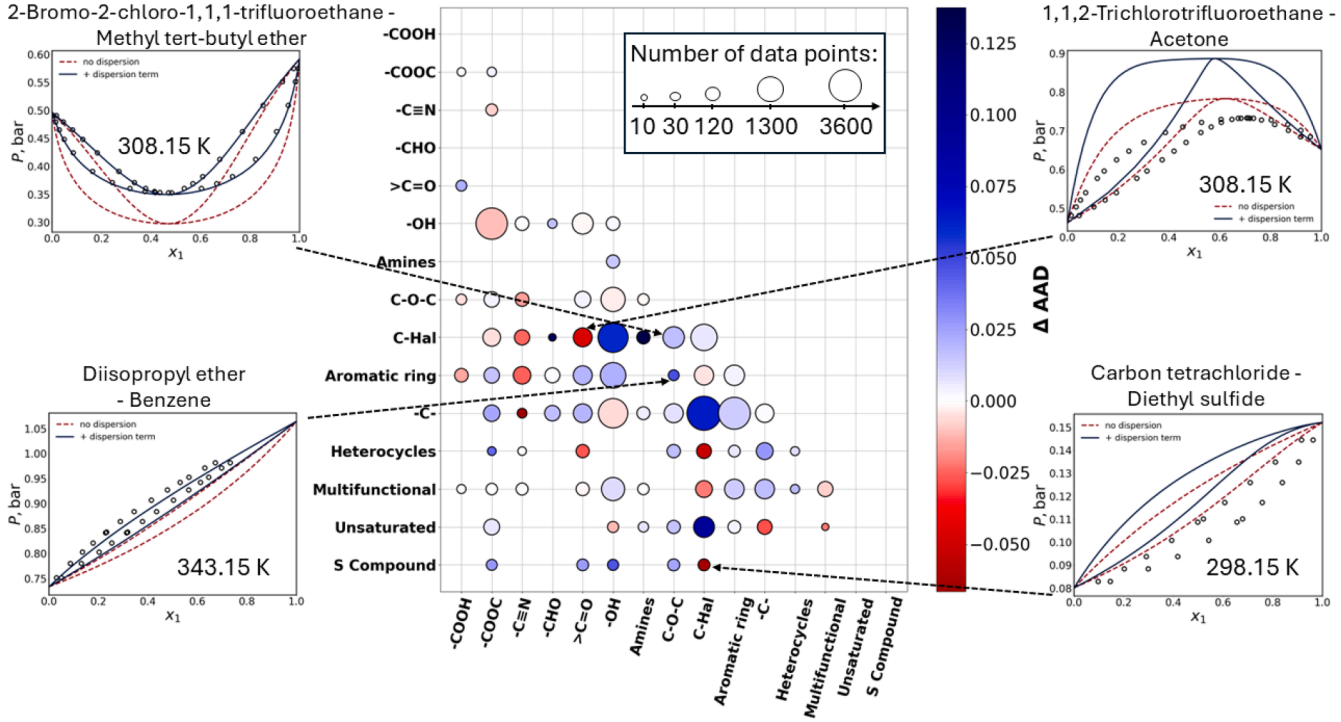

Fig. 11. Visualization of Δ AAD (Eq. (26)) calculated for model Eq. (24) + w_5 (Eq. (11)) for binary VLE data, accompanied by *P-xy* phase diagram examples. The colors in the phase diagrams indicate the order of magnitude of AAD, as shown by the color bar. Rows and columns represent the primary chemical functional groups of the components, while the circle sizes reflect the proportion of data points corresponding to molecules within each category. This plot concept is inspired by Fingerhut et al. (2017).

atomic parameters. An interesting observation pertains to the w_1 model, which indicates poor predictions for purely iodinated compounds but excels in predicting datasets containing both F and I. This behavior stems from the nature of the scaling itself. Since w_1 scales α_i by the number of segments assigned to each atom, the large size of I - atom results in a low weight for iodinated alkanes. This likely causes an underestimation of dispersive interactions within this group. Conversely, for datasets with both F and I, containing many highly fluorinated compounds, this scaling proves beneficial, while other models overestimate attractive interactions and, consequently, activity coefficients in these mixtures.

In conclusion, the proposed models achieve significant improvements in prediction accuracy compared to both the original model and our prior FH_6 model for halocarbons. In the following sections, we will extend these models to the complete dataset for broader validation.

3.2. The extended database

In the present study, we extended the halocarbon database (Table 1) by incorporating predominantly oxygen-containing organic compounds, as well as those containing nitrogen, sulfur, and phosphorus. The updated dataset encompasses a wide range of organic compound classes, and its chemical diversity provides a solid foundation for the development of a general model. Additionally, as summarized in Table 8, which includes both the new data and data from Table 1, the entire data collection consists not only of the predominant VLE, LLE, and IDAC data, but also partitioning coefficients and solvation free energies. These data types, commonly used for the parametrization and validation of thermodynamic models, capture distinct aspects of intermolecular interactions and offer valuable insights into the properties of liquid phases. The

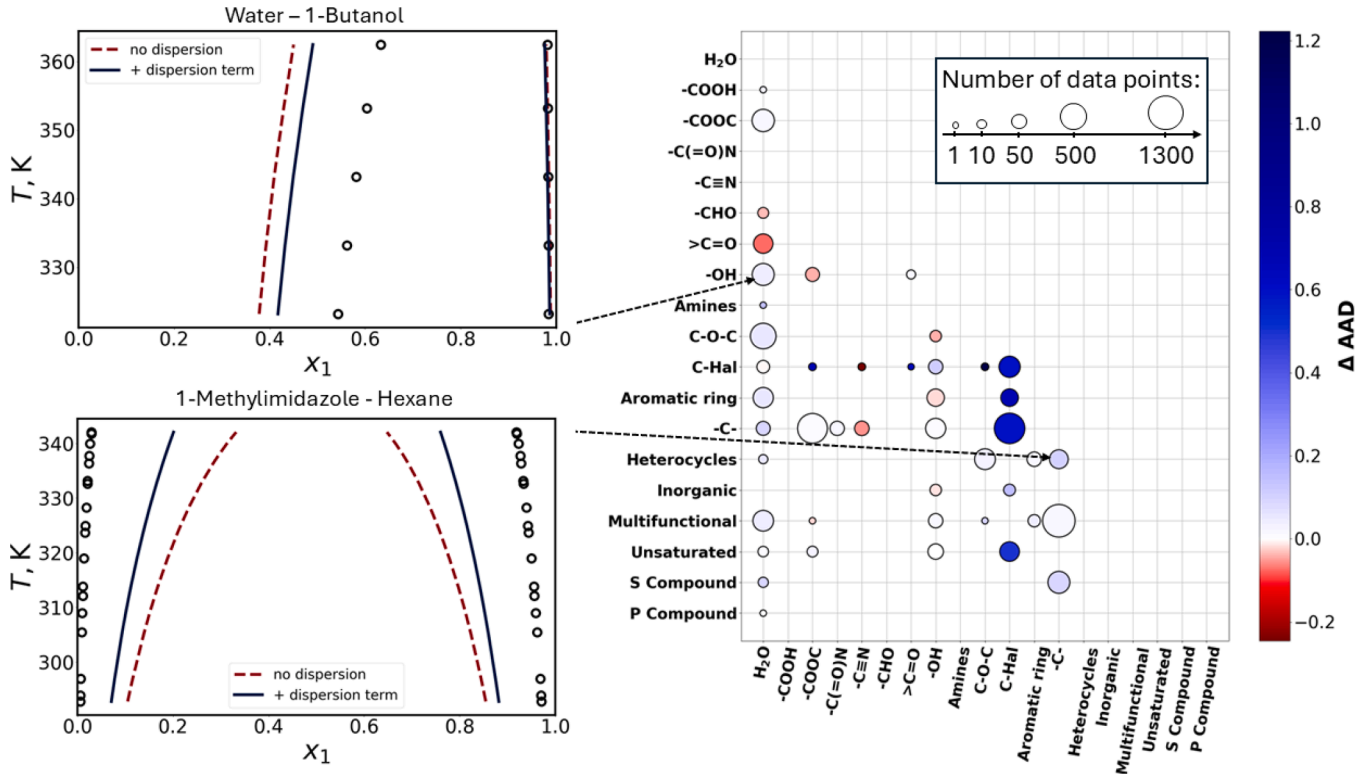

Fig. 12. Visualization of Δ AAD (Eq. (26)) calculated for model Eq. (24) + w_5 (Eq. (11)) for binary LLE data, accompanied by LLE phase diagram examples. The colors in the phase diagrams indicate the order of magnitude of AAD, as shown by the color bar. Rows and columns represent the main chemical functional groups of the components. The size of the circles indicates the proportion of data points corresponding to molecules within each category. The concept for this plot is inspired by Fingerhut et al. (2017).

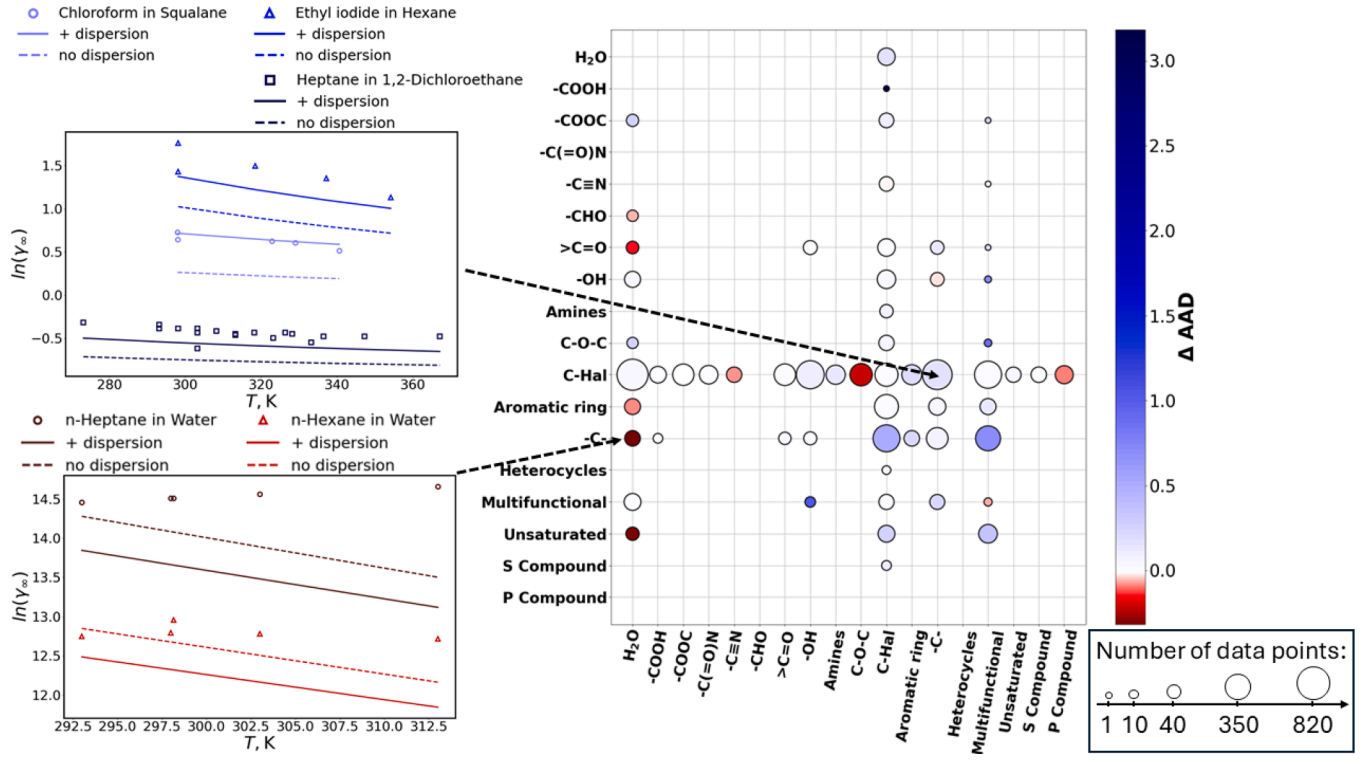

Fig. 13. Visualization of Δ AAD (Eq. (26)) calculated for model Eq. (24) + w_5 (Eq. (11)) for binary IDAC data, accompanied by its temperature dependence examples, colored according to the order of magnitude of AAD, as shown by the color bar. Rows (solutes) and columns (solvents) represent the main chemical functional groups of the components. The size of the circles indicates the proportion of data points corresponding to molecules within each category. The concept for this plot is inspired by Fingerhut et al. (2017).

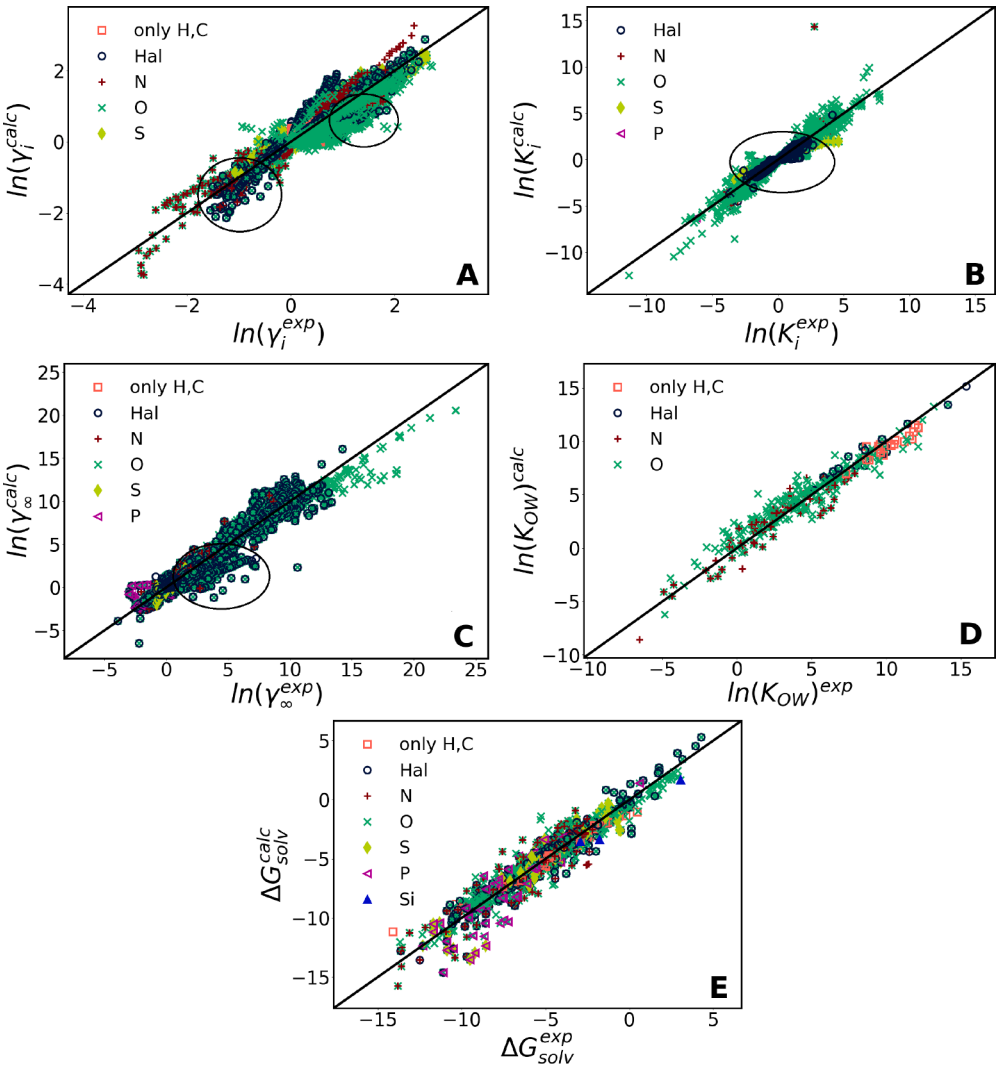

Fig. 14. Parity plots comparing predicted values (model Eq. (24) + w_5 (Eq. (11)) with experimental data across various datasets: VLE activity coefficients (**A**), LLE partition coefficients (**B**), IDAC (**C**), $K_i^{org/aq}$ (**D**), and ΔG_{solv} (**E**). The symbols are colored according to the type of atoms in the data. Regions with the most significant improvements are highlighted with circles.

majority of the data was obtained from various open-source databases. Specifically, most of the VLE data was extracted from the KDB database (KDB, 2024), rigorously tested for thermodynamic consistency, and used to calculate activity coefficients when it met the consistency criteria defined by the Van Ness test. Details regarding the sources and processing of the partitioning coefficients, IDAC, and solvation free energy data are available in Gerlach et al. (2022), Müller et al. (2025). To extend our LLE data collection, we utilized the open-source NIST database (Riccardi et al., 2021), extracting the required information from the JSON/XML files provided.

The general model parameters, without both the misfit correction for the F-atom and the dispersion term, were initially re-optimized using the entire dataset. These parameters (a_{eff}, α_{mf}, c_{hb}, σ_{hb}, r_{av}, f_{corr}, c_{hb}^T) pertain solely to the residual term, as the FH combinatorial term does not require additional parameters.

Following this, the models that demonstrated the best performance for halocarbons (Table 7) were extended to the complete dataset. We will further refer to the implementation including these models as openCOSMO-RS 25a. Subsequently, the same general parameters, along with E_{corr}^F and m_{vdW} were optimized. This simultaneous regression was performed due to the coupling between these parameters; the introduction of the dispersion term inevitably influences both the misfit and hydrogen-bonding terms. Notably, the dispersion effects, which were previously approximated within these terms, are now explicitly accounted for. Given the computational demands of these full-data optimizations, they were executed on a high-performance computing cluster. Depending on the number of cores (2–4), each optimization required 2–5 days of computational time, utilizing conventional algorithms like Nelder-Mead from the SciPy Python package (Virtanen et al., 2020). Table 9 summarizes the results of these regressions. The hydrogen-bonding parameters c_{hb} for models incorporating dispersion were consistently lower than for the model without dispersion. This reduction indicates that some of the attractive interactions previously attributed to hydrogen bonding are now accounted for by the dispersion term. For solvation free energy, the related parameters were independently optimized for each model, with the results provided in Table 10.

To provide a more detailed analysis of the modified model, we compared the AAD of the overall best-performing model from Table 9 (w_5 + Eq. (24)) with the unmodified version, labeled as 'no dispersion' in the same table. This comparison was conducted across various chemical groups by calculating the difference between those two models as

follows:

$$\Delta AAD = AAD(\text{no dispersion}) - AAD(+\ \text{dispersion}). \quad (26)$$

The results of this comparison for VLE are illustrated in Fig. 11. Positive values on the color bar indicate improved AADs, whereas negative values denote deteriorated AADs. Overall, approximately 70 % of the data groups showed improvement, with the most significant, as expected, in mixtures of halocarbons and hydrocarbons. However, this trend does not extend to all halocarbon mixtures, as only about half of the groups improved. For instance, systems involving halocarbons and ethers consistently demonstrated better accuracy, whereas mixtures of halocarbons with ketones or sulfur-containing compounds performed worse. Representative *P-xy* phase diagrams illustrating these trends are shown in Fig. 11. Notably, the two *P-xy* phase diagrams on the right-hand side of Fig. 11 suggest a tendency to overestimate attractive interactions. This finding further highlights the potential benefits of adopting a more sophisticated combining approach. Challenges also persist in modeling mixtures of halocarbons with compounds containing unsaturated carbon atom and heteroatoms. In particular, polar oxygenated mixtures likely require additional refinement, potentially through integration with a hydrogen bonding term to ensure a more balanced representation of different contributions.

Despite these remaining challenges, we consider the modified model an improvement for VLE predictions, as the majority of data groups showed enhanced performance. Furthermore, these improvements are not limited to halogenated systems, as demonstrated by the example of the diisopropyl ether–benzene system in Fig. 11.

A detailed analysis for LLE systems is presented in Fig. 12, which also includes two examples of LLE phase diagrams for improved categories. Deteriorated cases are not illustrated, but often, this simply indicates the absence of a phase split. Notably, most aqueous mixtures exhibit improved performance. Furthermore, nearly all halocarbon-containing groups improved, regardless of the second compound's chemical nature. However, challenges remain in modeling mixtures with polar oxygenated compounds. Despite these limitations, approximately 80 % of the data groups showed improved AADs. Moreover, the regression procedure applied to the LLE data, which is based on partition coefficients rather than on iterative phase split calculations, appears to yield reliable results. Combined with the proposed model modifications, this approach contributes to the overall accuracy of LLE predictions.

Finally, a similar AAD analysis was conducted for the IDAC data. As illustrated in Fig. 13, more than 80 % of the data categories showed improvement. In particular, most halocarbon systems exhibited significant decrease in AADs, both when halocarbons acted as solutes and as solvents. Since the first part of this work included relatively few halocarbon IDAC data points, we present additional examples of the temperature dependence of halocarbon–hydrocarbon systems in Fig. 13. However, challenges remain for mixtures where water serves as the solvent. For instance, in the cases of heptane and hexane in water, the original model outperforms the revised approach. This further underscores the necessity of future work on the hydrogen bonding term in conjunction with the dispersion term.

For highly polar systems, the incorporation of the dispersion term does not lead to profound improvements, as dispersive interactions play only a minor role in the overall molecular interactions. This trend is evident in the AAD differences shown in Figs. 11–13 particularly for mixture groups containing alcohols and water. Furthermore, Table 9 shows only minor improvements in AADs for the partitioning coefficients of predominantly polar systems in our data collection. The accuracy of the modified models for solvation free energies remains nearly unchanged, possibly due to the higher sensitivity of the model to the numerous atomic parameters involved in Eq. (20). We do not provide a detailed group analysis for these two types of data, however the respective parity plots as well as ones for other data types, are illustrated in Fig. 14. The circled regions in Fig. 14 highlight areas corresponding to data with the most notable improvements.

4. Conclusions

This study introduces an improved predictive openCOSMO-RS framework through the development of the dispersion term based on atomic polarizability as a quantum chemical descriptor. The modified model can be applied to design of a wide range of chemical engineering processes. In particular, those involving halogenated organic molecules can significantly benefit from the improved model developed in this study. For instance, accurate modeling of thermophysical properties and phase equilibrium of fluorocarbons is crucial for both classical industrial domains such as refrigeration and more recent but rapidly emerging environmental engineering problems such as PFAS remediation.

A significant part of this work involved developing a workflow for calculating atomic polarizabilities using the quantum chemical package ORCA 6.0 and integrating these outputs into the openCOSMO-RS framework. Multiple strategies were evaluated, particularly for halocarbon systems, including polarizability projections onto molecular cavities, scaling, and combining polarizabilities to compute segment-segment dispersion energies. The most reliable results were achieved using polarizability projections, further corrected using atomic ionization potentials.

Compared to our earlier model, which relied on atomic dispersion parameters, the new approach demonstrates superior performance while requiring substantially fewer adjustable parameters. This suggests that the proposed methodology may offer a more physically grounded approach to dispersion interactions.

Lastly, applying the new model to a diverse dataset encompassing a wide range of chemical systems demonstrated consistent improvements across various thermodynamic data types. The latest source codes for openCOSMO-RS, along with the RDKit/ORCA-based workflow for quantum chemical calculations incorporating atomic polarizabilities, are available on GitHub: https://github.com/TUHH-TVT. Future work will focus on refining openCOSMO-RS 25a and exploring the additional molecular descriptors to enhance its predictive capabilities further.

CRediT authorship contribution statement

Daria Grigorash: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Simon Müller:** Writing – review & editing, Writing – original draft, Software, Resources, Methodology, Investigation, Conceptualization; **Esther Heid:** Software, Methodology, Investigation, Conceptualization; **Frank Neese:** Software, Methodology, Conceptualization; **Dimitrios G. Liakos:** Software, Investigation, Conceptualization; **Christoph Riplinger:** Writing – review & editing, Software, Conceptualization; **Miquel Garcia-Ratés:** Software, Methodology, Conceptualization; **Patrice Paricaud:** Writing – review & editing; **Erling H. Stenby:** Supervision, Resources, Project administration, Funding acquisition; **Irina Smirnova:** Resources, Project administration; **Wei Yan:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Data availability

Data will be made available on request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank Edgar Ivan Sanchez Medina for generously sharing the digitalized DECHEMA database (Gmehling et al., 2008) for IDACs of halogenated compounds.

References

Arrad, M., Thomsen, K., Müller, S., Smirnova, I., 2024. Thermodynamic modeling using extended UNIQUAC and COSMO-RS-ES models: case study of the cesium nitrate-water system over a large range of temperatures. Fluid Phase Equilib. 580, 114037. https://doi.org/10.1016/j.fluid.2024.114037

Eckert, F., Klamt, A., 2003. Prediction of halocarbon thermodynamics with COSMO-RS. Fluid Phase Equilib. 210 (1), 117–141. https://doi.org/10.1016/S0378-3812(03)00166-3

Fingerhut, R., Chen, W.L., Schedemann, A., Cordes, W., Rarey, J., Hsieh, C.M., Vrabec, J., Lin, S.T., 2017. Comprehensive assessment of COSMO-SAC models for predictions of fluid-Phase equilibria. Ind. Eng. Chem. Res. 56 (35), 9868–9884. https://doi.org/10.1021/acs.iecr.7b01360

Flory, P.J., 1942. Themodynamics of high polymer solutions. J. Chem. Phys. 10 (1), 51–61. https://doi.org/10.1063/1.1723621

Garcia-Ratés, M., Neese, F., 2020. Effect of the solute cavity on the solvation energy and its derivatives within the framework of the gaussian charge scheme. J. Comput. Chem. 41 (9), 922–939. https://doi.org/10.1002/jcc.26139

Gerlach, T., Müller, S., de Castilla, A.G., Smirnova, I., 2022. An open source COSMO-RS implementation and parameterization supporting the efficient implementation of multiple segment descriptors. Fluid Phase Equilib. 560. https://doi.org/10.1016/j.fluid.2022.113472

Gerlach, T., Müller, S., Smirnova, I., 2018. Development of a COSMO-RS based model for the calculation of phase equilibria in electrolyte systems. AlChE J. 64. https://doi.org/10.1002/aic.15875

Gmehling, J., Tiegs, D., Medina, A., Soares, M., Bastos, J., Alessi, P., Kikic, I., Schiller, M., Menke, J., 2008. Activity Coefficients at Infinite Dilution. Vol. IX of DECHEMA Chemistry Data Series. DECHEMA.

González de Castilla, A., Müller, S., Smirnova, I., 2022. On the analogy between the restricted primitive model and capacitor circuits. Part II: a generalized Gibbs–Duhem consistent extension of the pitzer-debye-hückel term with corrections for low and variable relative permittivity. J. Mol. Liq. 360, 119398. https://doi.org/10.1016/j.molliq.2022.119398

González de Castilla, A., Müller, S., Smirnova, I., 2021. On the analogy between the restricted primitive model and capacitor circuits: semi-empirical alternatives for over- and underscreening in the calculation of mean ionic activity coefficients. J. Mol. Liq. 326. https://doi.org/10.1016/j.molliq.2020.115204

Gould, T., Bučko, T., 2016. C6 coefficients and dipole polarizabilities for all atoms and many ions in rows 1–6 of the periodic table. J. Chem. Theory Comput. 12 (8), 3603–3613. PMID: 27304856. https://doi.org/10.1021/acs.jctc.6b00361

Grensemann, H., Gmehling, J., 2005. Performance of a conductor-like screening model for real solvents model in comparison to classical group contribution methods. Ind. Eng. Chem. Res. 44 (5), 1610–1624. https://doi.org/10.1021/ie049139z

Grigorash, D., Müller, S., Paricaud, P., Stenby, E.H., Smirnova, I., Yan, W., 2024. A comprehensive approach to incorporating intermolecular dispersion into the openCOSMO-RS model. Part 1: halocarbons. https://arxiv.org/abs/2406.05244.

Heid, E., Fleck, M., Chatterjee, P., Schröder, C., MacKerell Alexander, , D.Jr., 2019. Toward prediction of electrostatic parameters for force fields that explicitly treat electronic polarization. J. Chem. Theory Comput. 15 (4), 2460–2469. PMID: 30811193. https://doi.org/10.1021/acs.jctc.8b01289

Heid, E., Szabadi, A., Schröder, C., 2018. Quantum mechanical determination of atomic polarizabilities of ionic liquids. Phys. Chem. Chem. Phys. 20, 10992–10996. https://doi.org/10.1039/C8CP01677A

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

Hsieh, C.-M., Sandler, S.I., Lin, S.-T., 2010. Improvements of COSMO-SAC for vapor-liquid and liquid–liquid equilibrium predictions. Fluid Phase Equilib. 297 (1), 90–97. https://doi.org/10.1016/j.fluid.2010.06.011

Hudson, G.H., McCoubrey, J.C., 1960. Intermolecular forces between unlike molecules. A more complete form of the combining rules. Trans. Faraday Soc. 56. https://doi.org/10.1039/tf9605600761

Jin, P., Murray, J.S., Politzer, P., 2004. Local ionization energy and local polarizability. Int. J. Quantum Chem. 96 (4), 394–401. https://doi.org/10.1002/qua.10717

Jiřiště, L., Klajmon, M., 2022. Predicting the thermodynamics of ionic liquids: what to expect from PC-SAFT and COSMO-RS? J. Phys. Chem. B 126 (20), 3717–3736. https://doi.org/10.1021/acs.jpcb.2c00685

KDB, 2024. CHERIC, Chemical Engineering and Materials Research Information Center. https://www.cheric.org/research/kdb.

Klajmon, M., 2022. Purely predicting the pharmaceutical solubility: what to expect from PC-SAFT and COSMO-RS? Mol. Pharm. 19 (11), 4212–4232. https://doi.org/10.1021/acs.molpharmaceut.2c00573

Klamt, A., 1995. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena starting from the question of why dielectric continuum models give a fairly good description of molecules. J. Phys. Chem. 99, 2224–2235. https://pubs.acs.org/sharingguidelines.

Klamt, A., 2003. Prediction of the mutual solubilities of hydrocarbons and water with COSMO-RS. Fluid Phase Equilib. 206 (1–2), 223–235. https://doi.org/10.1016/S0378-3812(02)00322-9

Klamt, A., Eckert, F., 2000. COSMO-RS: a novel and efficient method for the a priori prediction of thermophysical data of liquids. Fluid Phase Equilib. 172, 43–72. www.elsevier.nlrlocaterfluid.

Klamt, A., Eckert, F., 2007. Prediction, fine tuning, and temperature extrapolation of a vapor liquid equilibrium using COSMOtherm. Fluid Phase Equilib. 260 (2), 183–189. https://doi.org/10.1016/j.fluid.2007.07.055

Klamt, A., Eckert, F., Hornig, M., Beck, M.E., Brger, T., 2002. Prediction of aqueous solubility of drugs and pesticides with COSMO-RS. J. Comput. Chem. 23 (2), 275–281. https://doi.org/10.1002/jcc.1168

Klamt, A., Jonas, V., Bürger, T., Lohrenz, J. C.W., 1998. Refinement and parametrization of COSMO-RS. J. Phys. Chem. A 102 (26), 5074–5085. https://doi.org/10.1021/jp980017s

Kröger, L.C., Müller, S., Smirnova, I., Leonhard, K., 2020. Prediction of solvation free energies of ionic solutes in neutral solvents. J. Phys. Chem. A 124 (20), 4171–4181. PMID: 32336096. https://doi.org/10.1021/acs.jpca.0c01606

Krooshof, G. J.P., de With, G., 2024. Gibbs probability entropy and its implication to combinatorial entropy models. Fluid Phase Equilib. 584, 114146. https://doi.org/10.1016/j.fluid.2024.114146

Lin, S.T., Sandler, S.I., 2002. A priori phase equilibrium prediction from a segment contribution solvation model. Ind. Eng. Chem. Res. 41 (5), 899–913. https://doi.org/10.1021/ie001047w

London, F., 1937. The general theory of molecular forces. Trans. Faraday Soc. 33. https://doi.org/10.1039/tf937330008b

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

Morgado, P., Garcia, A.R., Martins, L. F.G., Ilharco, L.M., Filipe, E. J.M., 2017. Alkane coiling in perfluoroalkane solutions: a new primitive solvophobic effect. Langmuir 33 (42), 11429–11435. https://doi.org/10.1021/acs.langmuir.7b02516

Müller, S., González de Castilla, A., Taeschler, C., Klein, A., Smirnova, I., 2019. Evaluation and refinement of the novel predictive electrolyte model COSMO-RS-ES based on solid-liquid equilibria of salts and Gibbs free energies of transfer of ions. Fluid Phase Equilib. 483, 165–174. https://doi.org/10.1016/j.fluid.2018.10.023

Müller, S., González de Castilla, A., Taeschler, C., Klein, A., Smirnova, I., 2020. Calculation of thermodynamic equilibria with the predictive electrolyte model COSMO-RS-ES: improvements for low permittivity systems. Fluid Phase Equilib. 506. https://doi.org/10.1016/j.fluid.2019.112368

Murray, J.S., Seybold, P.G., Politzer, P., 2021. The many faces of fluorine: some non-covalent interactions of fluorine compounds. J. Chem. Thermodyn. 156, 106382. https://doi.org/10.1016/j.jct.2020.106382

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

National Center for Biotechnology Information, 2024. Ionization energy in the periodic table of elements. https://pubchem.ncbi.nlm.nih.gov/periodic-table/ionization-energy.

Neese, F., 2012. The ORCA program system. WIREs Comput. Mol. Sci. 2 (1), 73–78. https://doi.org/10.1002/wcms.81

Neese, F., Wennmohs, F., Becker, U., Riplinger, C., 2020. The ORCA quantum chemistry program package. J. Chem. Phys. 152 (22). https://doi.org/10.1063/5.0004608

Nguyen-Huynh, D., Passarello, J.-P., Tobaly, P., de Hemptinne, J.-C., 2008a. Modeling phase equilibria of asymmetric mixtures using a group-contribution SAFT (GC-SAFT) with a kij correlation method based on london's theory. 1. Application to CO_2 + n-alkane, methane + n-alkane, and ethane + n-alkane systems. Ind. Eng. Chem. Res. 47 (22), 8847–8858. https://doi.org/10.1021/ie071643r

Nguyen-Huynh, D., Tran, T.K.S., Tamouza, S., Passarello, J.-P., Tobaly, P., de Hemptinne, J.C., 2008b. Modeling phase equilibria of asymmetric mixtures using a group-contribution SAFT (GC-SAFT) with a kij correlation method based on london's theory. 2. application to binary mixtures containing aromatic hydrocarbons, n-alkanes, CO_2, N_2, and H_2S. Ind. Eni. Chem. Res. 47 (22), 8859–8868. https://doi.org/10.1021/ie071644j

Paes, F., Privat, R., Jaubert, J.-N., Sirjean, B., 2024. Predicting solvation energies of free radicals and their mixtures: a robust approach coupling the peng-robinson and COSMO-RS models. J. Mol. Liq. 401, 124641. https://doi.org/10.1016/j.molliq.2024.124641

Peng, D., Alhadid, A., Minceva, M., 2022. Assessment of COSMO-SAC predictions for solid-liquid equilibrium in binary eutectic systems. Ind. Eng. Chem. Res. 61 (35), 13256–13264. https://doi.org/10.1021/acs.iecr.2c00856

Prausnitz, J.M., Lichtenthaler, R.N., de Azevedo, E.G., 1999. Molecular Thermodynamics of Fluid-Phase Equilibria. Prentice-Hall PTR. third ed. edition.

Riccardi, D., Bazyleva, A., Paulechka, E., Diky, V., Magee, J.W., Kazakov, A.F., Townsend, S.A., Muzny, C.D., 2021. ThermoML/data archive. https://doi.org/10.18434/mds2-2422.

Ritter, E., Racheva, R., Storm, S., Müller, S., Ingram, T., Smirnova, I., 2016. Influence of inorganic salts on the phase equilibrium of triton x-114 aqueous two-phase systems. J. Chem. Eng. Data 61 (4), 1496–1501. https://doi.org/10.1021/acs.jced.5b00821

Rowlinson, J.S., Swinton, F.L., 1982. Chapter 5—Excess thermodynamic functions. In: Rowlinson, J.S., Swinton, F.L. (Eds.), Liquids and Liquid Mixtures third ed.. Butterworth-Heinemann. Butterworths Monographs in Chemistry, pp. 132–190. https://doi.org/10.1016/B978-0-408-24193-9.50009-6

Sachsenhauser, T., Rehfeldt, S., Klamt, A., Eckert, F., Klein, H., 2014. Consideration of dimerization for property prediction with COSMO-RS-DARE. Fluid Phase Equilib. 382, 89–99. https://www.sciencedirect.com/science/article/pii/S0378381214004877. https://doi.org/10.1016/j.fluid.2014.08.030

Saidi, C.N., Mielczarek, D.C., Paricaud, P., 2020. Predictions of solvation Gibbs free energies with COSMO-SAC approaches. Fluid Phase Equilib. 517. https://doi.org/10.1016/j.fluid.2020.112614

Scott, R.L., 1958. The anomalous behavior of fluorocarbon solutions. J. Phys. Chem. 62 (2), 136–145. https://doi.org/10.1021/j150560a002

Siebert, E.M.D., Knobler, C.M., 1971. Interaction virial coefficients in hydrocarbon-fluorocarbon mixtures. J. Phys. Chem. 75 (25), 3863–3870. https://doi.org/10.1021/j100694a014

Song, W., Rossky, P.J., Maroncelli, M., 2003. Modeling alkane + perfluoroalkane interactions using all-atom potentials: failure of the usual combining rules. J. Chem. Phys. 119 (17), 9145–9162. https://doi.org/10.1063/1.1610435

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A. R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods 17, 261–272. https://doi.org/10.1038/s41592-019-0686-2