Estimation of the binary interaction parameter kij of the PC-SAFT Equation of State based on pure component parameters using a QSPR method

Stavrou, M; Bardow, A; Gross, J

DOI
10.1016/j.fluid.2015.12.016

Publication date
2016

Document Version
Final published version

Published in
Fluid Phase Equilibria

Fluid Phase Equilibria 416 (2016) 138–149

Contents lists available at ScienceDirect

Fluid Phase Equilibria

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

Estimation of the binary interaction parameter k_{ij} of the PC-SAFT Equation of State based on pure component parameters using a QSPR method

Marina Stavrou [a, b], André Bardow [c], Joachim Gross [a, *]

[a] Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, Pfaffenwaldring 9, 70569, Stuttgart, Germany
[b] Process and Energy Laboratory, Delft University of Technology, Leeghwaterstraat 39, 2628, CB Delft, The Netherlands
[c] Institute of Technical Thermodynamics, RWTH Aachen University, Schinkelstrasse 8, 52062, Aachen, Germany

ARTICLE INFO

Article history:
Received 21 October 2015
Received in revised form
6 December 2015
Accepted 9 December 2015
Available online 19 December 2015

Keywords:
Binary interaction parameter
PC-SAFT
QSPR
VLE
CAMD

ABSTRACT

Statistical Associating Fluid Theory (SAFT) equations of state (EoS) for mixtures require cross-interaction parameters. For real systems, combining rules, such as the Lorenz-Berthelot combining rules, have to be corrected using at least one binary interaction parameter, k_{ij}. Values of k_{ij} are usually adjusted to experimental data of phase equilibria. Here, we correlate k_{ij} to the pure component parameters of the Perturbed Chain – Statistical Associating Fluid Theory (PC-SAFT) EoS, using a Quantitative Structure Property Relationship (QSPR) model. The coefficients of the proposed QSPR model are regressed separately for mixtures with non-associating components and for mixtures with associating components. The QSPR model is validated using the statistical measures of the QSPR method. We compare the values of k_{ij} that are estimated from the QSPR model to values of k_{ij} estimated from London's dispersive theory. Phase equilibrium calculations carried out with these two approaches of estimating k_{ij} values are compared to experimental data. The estimation of k_{ij} values as function of the pure component PC-SAFT parameters can be applied to problems of process design and in Computer Aided Molecular Design (CAMD), to allow for calculations that are reasonably accurate and independent from the availability of experimental mixture data.

1. Introduction

Statistical Associating Fluid Theory (SAFT) equations of state (EoS) are expressions of the residual Helmholtz energy that represent approximate algebraic solutions of equations from statistical mechanics [1–4]. For mixtures, a SAFT EoS requires the intermolecular potential ϕ_{ij} between the unlike molecules i and j. Usually, a conformal solution theory is applied, requiring cross-interaction parameters, such as a size-parameter σ_{ij} and an energy parameter ε_{ij}. Strictly speaking the $i-j$ cross-interaction potential, just as the potential between like $i-i$ species, has to be determined by quantum mechanical calculations. For van der Waals (dispersive) interactions, however, combining rules are often successful for estimating of cross-interaction parameters. Most variants of the SAFT-type EoS use the Lorenz-Berthelot combining rules: $\sigma_{ij}=1/2(\sigma_i+\sigma_j)$ and $\varepsilon_{ij}=\varepsilon_i\varepsilon_j$. For an accurate correlation of real systems, the combining rules have to be corrected using at least one interaction parameter, k_{ij} for every binary pair. As a modification of the Lorenz-Berthelot combining rules, the binary interaction parameter k_{ij} is introduced as a correction to the dispersive energy parameter for the binary pair: $\varepsilon_{ij}=(1-k_{ij})\varepsilon_i\varepsilon_j$. Binary interaction parameters are often adjusted to experimental phase equilibrium data of the binary mixtures.

SAFT-type EoS were successfully applied to a wide range of complex systems, that include polar and associating components, polymers, ionic liquids, pharmaceuticals and bio-molecules (see for example ref. [1–3,5–8]). SAFT-type EoS are often more accurate and more predictive than cubic EoS and extrapolate more reliably than g^E-models [1]. For mixtures of substances where the intermolecular potentials are dominated by van der Waals interactions, SAFT-type EoS often give accurate results, even without binary correction ($k_{ij}\simeq 0$). Many mixtures of industrial interest however require adjusting a binary interaction parameter k_{ij}.

Experimental data of phase equilibria are not always available

* Corresponding author.
E-mail address: gross@itt.uni-stuttgart.de (J. Gross).

http://dx.doi.org/10.1016/j.fluid.2015.12.016
0378-3812/

for the estimation of k_{ij} values. Consider the design of novel processes or a molecular design problem. In Computer Aided Molecular Design (CAMD) for example, thermodynamic models are required to evaluate some objective function quantifying the performance of the designed molecule [9,10]. CAMD problems are formulated as reverse property prediction problems [11]. In CAMD problems it is not possible to adjust binary interaction parameters on experimental mixture data. Early stages of process design and CAMD rely on predictions of mixture behavior, which in our context means predictions of k_{ij} values.

O'Connell et al. [10] studied several thermodynamic property models and suggested SAFT-type EoS to have a potential for accurate property predictions in CAMD. In a recent study, Ng et al. [12] gave a thorough review of the various approaches and advances in CAMD methods. Among the most recent advances are CAMD frameworks that use a SAFT-type model [13–19]. Adjiman et al. [20] presented the rapid progress made in the field of SAFT-based CAMD and discussed the role of SAFT-type EoS in tackling previously challenging problems in molecular and process design. In a SAFT-based CAMD framework, variables that characterize the structure of the optimized molecule are utilized as additional degrees of freedom inside the process optimization. The molecular optimization can thereby be integrated with the process optimization problem, based on an objective function for the entire process. The problem of integrated process and fluid design has in the past been circumvented, by defining individual or collected property targets for the optimized fluid [21–23].

Pereira et al. [13,15] introduced the framework of Computer Aided Molecular and Process Design (CAMPD) that uses the SAFT-VR EoS [24]. In their work, the authors applied CAMPD for the solvent selection among alkane blends. The molecular search space was defined over the homologous series of n-alkanes and the number of carbon atoms of the n-alkane was treated as a continuous molecular optimization variable. The necessary SAFT-VR parameters were expressed as functions of the molecular mass and therefore as functions of the optimized number of carbon atoms. The values of k_{ij} for the SAFT-VR EoS were adjusted to experimental phase equilibrium data of binary mixtures relevant to the examined system and they were held constant throughout CAMPD. Recently, Burger et al. [25] proposed a CAMPD method that uses the SAFT-γ Mie GC EoS [26] and a hierarchical optimization approach. The authors demonstrated the proposed method for the problem of solvent selection adopted from Pereira et al. In their work, Burger et al. [25] applied a group contribution (GC) approach for SAFT and considered additional chemical families extending the search space to linear alkyl ethers.

The framework of Continuous Molecular Targeting - Computer Aided Molecular Design (CoMT-CAMD), established by Bardow et al. [14], uses the PC-SAFT EoS [27–30]. CoMT-CAMD considers the parameters of the thermodynamic model (e.g. representing a solvent) as additional degrees of freedom in an integrated process and fluid optimization problem. In CoMT-CAMD the discrete search space of parameters representing a real substance is relaxed to a continuous parameter domain. Specifically, the molecular optimization variables (PC-SAFT pure component parameters) are treated as continuous variables. This relaxation allows formulating a non-linear optimization problem with a single objective function, whereby a detailed process model can be maintained without the need for pre-selecting candidate molecules. CoMT-CAMD has been implemented, so far, for the selection and design of working fluids for organic rankine cycles (ORCs) [16,19] and for the selection and design of physical solvents for CO_2 capture [18,19]. In phase equilibrium calculations no binary interaction correction was introduced to the PC-SAFT EoS for binary systems involving the optimized fluid ($k_{ij} = 0$).

The result of a CAMD approach is determined by the employed thermodynamic model. The accuracy of the SAFT model therefore plays a crucial role in the predictive capability of the CAMD framework.

In order to improve predictions of mixture properties (with $k_{ij} = 0$), many studies have revisited the Lorentz-Berthelot combining rules [31–38]. More recently, Haslam et al. [39] derived a new combining rule using a generalization of the Hudson-McCoubrey combining rules, including additional terms for the dipole–dipole and the dipole-induced dipole interactions. The authors presented results of their theory with the SAFT-VR EoS [24] for mixtures with large non-polar and polar components and for mixtures with a single associating component. Singh et al. [40] and Leonhard et al. [41] proposed the prediction of pure component parameters based on quantum mechanical calculations and a new combining rule for the unlike-dispersion interaction, derived from London's dispersive theory. Leonhard et al. applied the proposed combining rule with the PC-SAFT EoS. Both of the above approaches have a theoretical background and gave good results for mixtures.

Group contribution approaches are promising for estimating binary interaction corrections k_{ij} to the Lorenz-Berthelot combining rules. Peters et al. [42] proposed a group contribution method for determining binary interaction parameters of the PC-SAFT EoS for polymer-solvent mixtures. Inspired by London's dispersive theory, Huynh et al. [43,44] proposed the correlation of k_{ij} to pseudo-ionization energies of the constituents of the mixture. The method was developed for the GC-SAFT EoS [45] and was shown to lead to very good results. London's dispersive theory applied to the Mie potential function has been used by Coutinho et al. [46] to derive a new combining rule for the cross-energy parameter in cubic EoS and in the Cubic Plus Association EoS [47].

Due to the analogy between the combining rules of SAFT-type EoS and the combining rules in cubic EoS, we mention the work of Shacham et al. [48] and the work of Abudour et al. [49]. Both studies proposed the prediction of k_{ij} values for cubic EoS using a Quantitative Structure Property Relationship (QSPR) method with automatically generated molecular descriptors.

The CoMT-CAMD approach requires the prediction of mixture properties (i.e. estimating a suitable k_{ij}) based on pure component parameters of PC-SAFT for the optimized fluid. On that grounds, two of the approaches for estimating k_{ij} mentioned above particularly lend themselves for the CoMT-CAMD framework: the estimation of k_{ij} based on predicted ionization potentials and the QSPR estimation method.

In this work, we relate the binary interaction parameter k_{ij} to pure component parameters of both substances, in order to allow for predicting mixture properties in CAMD applications, especially in the CoMT-CAMD framework. Two approaches are examined. Values of k_{ij} are estimated based on London's dispersive theory using experimental ionization potentials. Further, we propose a multilinear regression model for estimating k_{ij} based only on the PC-SAFT pure component parameters. For the development of the proposed model, a QSPR method is applied.

2. Estimation of k_{ij} based on London's dispersive theory

For a mixture of conformal fluids, i.e. fluids with the same functional form of van der Waals intermolecular pair potentials, theoretical expressions of the k_{ij} value can be derived from London's dispersive theory. Most common is the expression based on the work of Hudson and McCoubrey [50]. Hudson and McCoubrey wrote the equation of London's attractive potential ϕ_{ij}^{disp} as function of the ionization potentials I_i and I_j and the (scalar valued) static polarizabilities α_i and α_j of the two constituents of the mixture

$$\phi_{ij}^{\text{disp}} = -\frac{3}{2}\frac{I_i I_j}{(I_i + I_j)}\frac{\alpha_i \alpha_j}{r_{ij}^6} \tag{1}$$

Eq. (1) combined with the attractive part of the Lennard-Jones potential

$$\phi_{ij}^{\text{LJ}} = 4\varepsilon_{ij}\left(\left[\frac{\sigma_{ij}}{r_{ij}}\right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}}\right]^{6}\right) \tag{2}$$

and using the arithmetic mean for the segment diameter parameter σ_{ij}

$$\sigma_{ij} = \frac{1}{2}(\sigma_i + \sigma_j) \tag{3}$$

led to an expression for the depth of the attractive potential well due to dispersive interactions ε_{ij}, i.e. the Hudson-McCoubrey rules

$$\varepsilon_{ij} = \left[2.\frac{(I_i I_j)^{1/2}}{(I_i + I_j)}\right]\cdot\left[2^6.\frac{\sigma_i^3\sigma_j^3}{(\sigma_i + \sigma_j)^6}\right]\cdot\sqrt{\varepsilon_i\varepsilon_j}. \tag{4}$$

For a SAFT-type EoS Eq. (4) leads to an approximation of k_{ij} as function of the segment diameter parameters and the ionization potentials of the two Lennard-Jones fluids in the mixture

$$k_{ij} = 1 - \left[2.\frac{(I_i I_j)^{1/2}}{(I_i + I_j)}\right]\cdot\left[2^6.\frac{\sigma_i^3\sigma_j^3}{(\sigma_i + \sigma_j)^6}\right]. \tag{5}$$

For details on the derivation of Eq. (4) and Eq. (5) we refer to the original work of Hudson and McCoubrey [50] and to the study of Haslam et al. [39].

The expression in Eq. (5), derived from London's theory, accounts only for the asymmetry in dispersive interactions. For mixtures of polar or associating components, however, k_{ij} does not only serve as correction to the dispersive intermolecular potential, as already pointed out by Hiza and Duncan [32] and Kontogeorgis [51]. Rather, k_{ij} corrects a SAFT-type model regarding any other model deficiency, including those arising from attractive interactions, not explicitly accounted for. However, London's dispersive theory is appealing for (CoMT-) CAMD applications, and more generally for predicting mixture properties, because only pure component properties of the mixtures' constituents are required. In this work, we evaluate phase equilibria predicted using k_{ij} values from Eq. (5) applied with the PC-SAFT model, using experimental values for the required ionization potentials.

3. Multivariate regression model for k_{ij} prediction

3.1. Contribution of asymmetric intermolecular potentials to the value of k_{ij}

QSPR studies focus mainly on the prediction of properties for pure substances. In most QSPR studies, QSPR software like CODESSA [52] or DRAGON [53] is employed to generate constitutional, topological, geometrical, electrostatic, quantum chemical or thermodynamic molecular descriptors (e.g. Refs. [49,54,55]). The significant descriptors are then selected using various stochastic or deterministic methods. Subsequently, the selected descriptors are combined to generate QSPR models. The resulting QSPR models are evaluated and compared based on multivariate statistics. In this study, we aim to develop a QSPR model for the estimation of k_{ij}, whereby k_{ij} is a mixture attribute. While a large number of descriptors are available for pure substances, the number of descriptors characterizing molecular pairs is very limited [56]. Here, we define the QSPR descriptors in an ad hoc manner, based on two principles. The descriptors for the estimation of k_{ij} should: First, be a function of the PC-SAFT pure component parameters only, and second, effectively relate dissimilarities in the molecular structure of two components to the k_{ij} value of their binary mixture.

In PC-SAFT, each pure component is identified by a unique set of molecular parameters: the segment number m, the segment size parameter σ, the dispersive energy parameter ε/k, the dipole moment μ, the quadrupole moment Q, the association energy parameter ε^{AB}/k and the effective association volume κ^{AB}. The equation of the QSPR model for the estimation of the value of k_{ij} (k_{ij}^{QSPR}) as function of the molecular parameters of the mixture components is therefore

$$k_{ij}^{\text{QSPR}} = \sum_{L=1}^{N_d} c_L \cdot D_L\left(\overline{p}_i, \overline{p}_j\right) \tag{6}$$

where N_d is the number of descriptors, $D_L(\overline{p}_i, \overline{p}_j)$ are the descriptors as function of the PC-SAFT molecular parameters $\overline{p}_i = \{m_i, \sigma_i, \varepsilon_i/k, \mu_i, Q_i, \varepsilon_i^{AB}/k\}$ and $\overline{p}_j = \{m_j, \sigma_j, \varepsilon_j/k, \mu_j, Q_j, \varepsilon_j^{AB}/k\}$ of the two components, i and j, of the mixture, and c_L are the corresponding regression coefficients.

Let $p_{k,i}$ be the k^{th} element of the PC-SAFT pure component parameter vector $\overline{p}_i$, e.g. $p_{1,i} = m_i$. Relations such as ratios of the pure component parameters, $a_{ij} = p_{k,i}/p_{k,j}$ or absolute differences, $\delta_{ij} = \left|p_{k,i} - p_{k,j}\right|$ can be used as measures of the asymmetry of intermolecular potentials in the binary mixture. Parameter ratios a_{ij} and absolute differences δ_{ij} can also be defined over combinations of the PC-SAFT parameters: $a_{ij} = h(\overline{p}_i)/h(\overline{p}_j)$ and $\delta_{ij} = \left|h(\overline{p}_i) - h(\overline{p}_j)\right|$. The departure of a_{ij} from unity and of δ_{ij} from zero quantify the difference between the $i-i$ and the $j-j$ property for the two components of the mixture, respectively. In order to make the parameter ratio a_{ij} invariant for interchanging the component indices i and j for the pair of substances, we further define the ratio operator

$$\left\langle a_{ij}\right\rangle = \begin{cases} a_{ij}^{\frac{(a_{ij}-1)}{\left|a_{ij}-1\right|}} & a_{ii} \neq a_{jj} \\ 1 & a_{ii} = a_{jj} \end{cases} \tag{7}$$

Measures of this type were combined to form the candidate descriptors for the QSPR model.

Unlike intermolecular potentials are caused by asymmetric dispersive, polar and associating forces. In order to quantify the contribution of the aforementioned asymmetries to the k_{ij} value, we defined eight descriptors. The initial form of the descriptors was mainly motivated by parameter combinations as they appear in the mathematical formulation of the PC-SAFT EoS. The exponents used in the final mathematical formulation of the descriptors are defined empirically.

3.1.1. Dispersive interactions

We define the descriptor D^{LJ} to express the asymmetry in the dispersive intermolecular potential, as

$$D^{\text{LJ}} = 1 - \left\langle \frac{\sigma_i^3(\varepsilon_i/k)^2}{\sigma_j^3(\varepsilon_j/k)^2}\right\rangle \tag{8}$$

by taking into account the segment diameter σ_i,σ_j and the Lennard-Jones energy potentials ε_i/k,ε_j/k. Our starting point for this descriptor has been the ratio $\langle\frac{m_i^2\sigma_i^3(\varepsilon_i/k)^2}{m_j^2\sigma_j^3(\varepsilon_j/k)^2}\rangle$ as motivated from the first order term of the perturbation theory for dispersive attraction. The exponents were varied empirically, usually in the integer-range of

±2. D^{LJ} will approach zero for very similar components and if no other types of interactions exist (e.g. polar or associating interactions).

3.1.2. Polar interactions

For contributions to the value of k_{ij} due to polar interactions, we define four descriptors. The contribution due to asymmetry in dipole–dipole interactions is expressed through the descriptors $D^{dd,a}$ and $D^{dd,b}$. Descriptor $D^{dd,a}$ measures the absolute difference in the reduced dipole moments.

$$D^{dd,a} = \left| \frac{\mu_i}{\sqrt{m_i \sigma_i^3 (\varepsilon_i/k)}} - \frac{\mu_j}{\sqrt{m_j \sigma_j^3 (\varepsilon_j/k)}} \right| \tag{9}$$

The difference in the reduced dipole moments contributes in a different way to the asymmetry in mixtures with only one dipolar component than in the case when both mixture components are dipolar. In order to distinguish these two cases, we introduce the descriptor $D^{dd,b}$.

$$D^{dd,b} = \sqrt{\mu_i \mu_j} \cdot (\mu_i - \mu_j)^2 \tag{10}$$

A further contribution due to asymmetry in quadrupole–quadrupole interactions is given by the descriptor D^{qq}. Descriptor D^{qq} considers the difference of the scaled quadrupole moments over the potential energy of the chain molecule. The difference in the quadrupole moments is scaled by the ratio of segment diameters according to

$$D^{qq} = \left[\frac{Q_i}{m_i (\varepsilon_i/k)} - \frac{Q_j}{m_j (\varepsilon_j/k)} \right]^2 \cdot \left\langle \frac{\sigma_i^5}{\sigma_j^5} \right\rangle \tag{11}$$

Finally, we account for the case that dipole–dipole and quadrupole–quadrupole interactions occur simultaneously with the descriptor D^{dq}.

$$D^{dq} = \left| \frac{\mu_i}{\sqrt{m_i \sigma_i^3 (\varepsilon_i/k)}} - \frac{\mu_j}{\sqrt{m_j \sigma_j^3 (\varepsilon_j/k)}} \right| \cdot \left[\frac{Q_i}{m_i (\varepsilon_i/k)} - \frac{Q_j}{m_j (\varepsilon_j/k)} \right]^2 \tag{12}$$

3.1.3. Association

Naive combining rules are not suited to describe cross-association of two substances, because cross-association can not be detached from electrostatic concepts of partial charge distributions in individual molecules. In that light, it is surprising that simple combining rules have shown promising results for a collection of chemical species [28,51,57]. (A group contribution approach on the other hand is in our view a rather promising concept). In studies where combining rules for cross-association are applied, a binary interaction correction is often applied to the dispersive interactions, rather than to the cross-association.

Consequently, mixtures with (cross-)associating substances are demanding for our approach, where the binary mixture shall be described based on pure component parameters of both species. With some reservation, we include associating mixtures in this study. It is important to note that the k_{ij} parameters adjusted to such mixtures usually show rather high (positive or negative) values, because of the uncertainty in cross-associating interactions. We account for the contribution of self- and cross-association effects with three descriptors. Descriptor $D^{assoc,s}$ measures the strength of self-association for component i against the strength of self-association for component j.

$$D^{assoc,s} = \sqrt{(\varepsilon_i/k)(\varepsilon_j/k)} \cdot \left[\sigma_i^3 \left(\varepsilon_i^{AB}/k \right) - \sigma_j^3 \left(\varepsilon_j^{AB}/k \right) \right]^2 \tag{13}$$

Kontogeorgis and Folas [58] distinguish between five different types of cross-association. We turn to the effect of cross-association between two dipolar components, when at least one of them is self-associating. Kleiner and Sadowski [57] proposed an approach to account for the cross-association that is likely to occur for such mixtures; they refer to these interactions as 'induced association'. We also observed that the combining rule for induced association [28,59].

$$\varepsilon^{A_i B_j} = \frac{1}{2} \left(\varepsilon_i^{AB} + \varepsilon_j^{AB} \right) \tag{14}$$

$$\kappa^{A_i B_j} = \sqrt{\kappa^{A_i B_i} \kappa^{A_j B_j}} \cdot \left(\frac{\sigma_i \sigma_j}{1/2(\sigma_i + \sigma_j)} \right)^3 \tag{15}$$

noticeably improves the accuracy of phase equilibrium calculations for this type of mixtures. The descriptor for the contribution due to induced association is defined as

$$D^{assoc,c} = \sqrt{\mu_i \mu_j} \cdot \frac{\left[\left(\varepsilon_i^{AB}/k \right) + \left(\varepsilon_j^{AB}/k \right) \right]^2}{\sqrt{\left(m_i \sigma_i^3 \cdot m_j \sigma_j^3 \right)}} \tag{16}$$

The indices s and c in Eqs. (13) and (16) are for "self" and "cross" association contributions, respectively. The contribution to the value of k_{ij} due to induced association accounts for the mutual strength of the dipole moments [60]. The contribution of the association energy parameters and the dipole moments in the descriptor $D^{assoc,c}$ is scaled by the molecular volume of the two components. This provides a better description for mixtures with small but strong associating molecules (e.g. mixtures of acetic acid) [61]. In mixtures with one associating and one non-polar, non-associating component, the effect of induced-association should not be accounted for. In this case, the descriptor $D^{assoc,c}$ is therefore equal to zero and no discrete decisions are necessary to distinguish between cases. The contribution of induced association in the asymmetry of a mixture is different when both components are self-associating species than when the mixture contains only one self-associating component and one dipolar, non-associating component. We introduced the third descriptor $D^{assoc,sc}$, in order to decouple the two cases when describing these types of mixtures with the same model simultaneously. The descriptor $D^{assoc,sc}$ is active only for mixtures with two self-associating components

$$D^{assoc,sc} = (\mu_i \mu_j) \cdot \sqrt{\left(\varepsilon_i^{AB}/k \right) \left(\varepsilon_j^{AB}/k \right)} \cdot \left| \left[\sigma_i^3 \left(\varepsilon_i^{AB}/k \right) \right]^2 - \left[\sigma_j^3 \left(\varepsilon_j^{AB}/k \right) \right]^2 \right|. \tag{17}$$

3.2. Pure component parameters

The necessary PC-SAFT pure component parameters were either adopted from Refs. [27–30,62] or, if not available, they were identified in the present work. In those cases, the pure component parameters $(m, \sigma, \varepsilon/k, \varepsilon^{AB}/k)$ were adjusted to experimental data of vapor pressure and liquid density. We have implemented the 2B association scheme [63] for all associating components. The association energy parameter ε^{AB}/k and the effective association volume

κ^{AB} are known to be strongly correlated, which is why a constant effective association volume $\kappa^{AB} = 0.03$ was used, as previously suggested by Ruether and Sadowski [64]. Dipole moments were taken from the DIPPR database [65] and they originate either from *ab initio* calculations or from measurements. The quadrupole moments were taken from Ref. [29].

3.3. Database of k_{ij} values adjusted to experimental data of phase equilibria

For isolating the effect of the different descriptors, we define four classes of substances: non-associating, non-polar (nAnP), non-associating, dipolar (nAdP), non-associating, quadrupolar (nAqP) and associating, dipolar (AdP). From the four classes of pure components, we define 10 groups of binary mixtures. The 10 groups of binary systems considered in our study and the number of mixtures per group are listed in Table 1.

For every binary mixture in our database, we adjusted the binary interaction parameter k_{ij} using experimental vapor–liquid equilibrium (VLE) data. The selected experimental data comprised sets of isothermal (P,x) and (P,x,y) data obtained from the Dortmund Database (DDB) [66], which passed the standard thermodynamic consistency tests, namely the point-to-point test and the area-test [67]. For mixtures with several experimental isotherms, we adjusted a single temperature-independent value of k_{ij}.

The objective function for adjusting the k_{ij} values is formulated on the combined residuals of the mole fraction of the liquid phase and the residuals of the pressure for every experimental point [68]:

$$F = \left[\sum_{l=1}^{n_{\text{exp}}} \frac{(\Delta x_l)^2 \cdot \left(\Delta P_l^{rel}\right)^2}{(\Delta x_l)^2 + \left(\Delta P_l^{rel}\right)^2} \right]^{1/2} \tag{18}$$

with n_{exp} as the number of experimental data points, $\Delta x_l = \left| x_l^{calc} - x_l^{exp} \right|$ the absolute residuals in the mole fraction of the liquid phase resulting from isobaric-isothermal flash calculations and $\Delta P_l^{rel} = \left| \frac{P_l^{calc}}{P_l^{exp}} - 1 \right|$ the relative residuals in pressure resulting from bubble point calculations.

3.4. Quantitative Structure Property Relationship (QSPR) for predicting k_{ij}

3.4.1. Multivariate regression

The QSPR model for predicting k_{ij} was built in steps: Initially, the QSPR model was developed for the simplest case; for predicting k_{ij} in mixtures of two non-associating, non-polar components (group 1), using only the descriptor for the asymmetry in the dispersive intermolecular potential D^{LJ}. Additional groups of binary mixtures were considered progressively and the model was supplemented with the necessary descriptors. In this way, the model was extended to all groups of mixtures with non-associating components (groups 1 to 6). Subsequently, the mixtures of associating components (groups 7 to 10) were considered. The QSPR model for all groups of binary mixtures reads

$$k_{ij}^{\text{QSPR}} = \sum_{L=1}^{N_d} c_L \cdot D_L \tag{19}$$

with model coefficients

$$c_L \in \left\{ c^{\text{LJ}}, c^{\text{dd,a}}, c^{\text{dd,b}}, c^{\text{qq}}, c^{\text{dq}}, c^{\text{assoc,s}}, c^{\text{assoc,c}}, c^{\text{assoc,sc}} \right\}$$

and with corresponding descriptors D_L

$$D_L \in \left\{ D^{\text{LJ}}, D^{\text{dd,a}}, D^{\text{dd,b}}, D^{\text{qq}}, D^{\text{dq}}, D^{\text{assoc,s}}, D^{\text{assoc,c}}, D^{\text{assoc,sc}} \right\}$$

as introduced in Eqs. (8)–(17). A non-weighted non-linear least squares problem is solved for the regression of the model coefficients c_L. The objective function for the model regression is defined as the total sum of the squared residuals between the values of k_{ij} individually adjusted to experimental data k_{ij}^{fit} and the values of k_{ij} calculated from the QSPR model k_{ij}^{QSPR} in Eq. (19) as

$$\min_{\bar{c}} Q = \sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}} \right)^2 \tag{20}$$

with N_{tr} the number of mixtures used for the regression. The least squares minimization was conducted using the solver *nlinfit* provided in Matlab [69]. For the 95% confidence interval of the model coefficients $\Delta c_{L,95}$ we use the *t*-distribution with $N_{\text{tr}} - N_d - 1$ degrees of freedom [70].

$$\Delta c_{L,95} = t_{0.95,N_{\text{tr}}-N_d-1} \cdot \sqrt{S^2} \cdot \sqrt{V_{LL}} \tag{21}$$

with S^2 as the estimated model variance

$$S^2 = \frac{\sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}} \right)^2}{N_{\text{tr}} - N_d - 1} \tag{22}$$

and with V_{LL} as the corresponding diagonal element of the coefficient covariance matrix.

3.4.2. Training and test set

The quality and the morphology of the data sets used to derive and validate the QSPR model coefficients (QSPR database) is decisive. The set used to derive the QSPR model coefficients (training set) should span the whole region of the descriptor space, it should be diverse and it should include data points close to the data points used for the external validation of the model (test set) [71,72]. In order to ensure these conditions, we define the training and test set

Table 1
The database of binary mixtures is divided in 10 groups (here denoted as G1 to G10). The binary mixtures are categorized according to the type of components they consist of: (a) non-associating, non-polar (nAnP), (b) non-associating, dipolar (nAdP), (c) non-associating, quadrupolar (nAqP) and (d) associating, dipolar (AdP). The number of mixtures for each subgroup in the database is listed.

		Comp. *j*			
		nAnP	nAdP	nAqP	AdP
Comp. *i*	nAnP	$^{(G1)}48$			
	nAdP	$^{(G2)}47$	$^{(G6)}32$		
	nAqP	$^{(G3)}44$	$^{(G4)}32$	$^{(G5)}5$	
	AdP	$^{(G7)}67$	$^{(G8)}90$	$^{(G9)}43$	$^{(G10)}53$

respectively in a three-step procedure: In step 1, we *a priori* detect and remove from the QSPR database the binary mixtures with unreliable k_{ij} values [70,73]. In step 2, we divide the QSPR database in training and test set, using a uniform design method for subset selection [74]. In step 3, we iteratively remove the outliers from the training set. The steps are discussed in more detail in the following.

In step 1 we use as measure for the quality of the adjusted k_{ij} values the value of the objective function F in Eq. (18), divided by the total number of the experimental data points, according to

$$f = F/n_{\text{exp}}. \tag{23}$$

The value of k_{ij} for a mixture m is considered as unreliable if $f_m > 0.95 \cdot \max(\bar{f})$ and the absolute average residuals AAD-x and AAD-P^{rel} are higher than 15%. Those cases either correspond to binary mixtures with experimental data of poor quality or to mixtures, where PC-SAFT leads to inaccurate results with the chosen pure component parameters.

After unreliable data were removed, the curated database (N_m binary mixtures) is divided into the training and the test set in step 2. In our case, we define the split fraction between the training and the test set $N_{\text{tr}}/N_{\text{ts}}$ equal to 9 (with $N_{\text{tr}} + N_{\text{ts}} = N_m$). For the selection of the training set, we have used an in-house implementation of the Kennard and Stone algorithm. The Kennard and Stone (KS) algorithm [75] is a well-established method for uniform design in the field of chemometrics [74]. The selection principle of the algorithm [75] ensures that the points which are excluded (test set) are close to the points of the design set (training set). Here, we implemented the KS algorithm individually for each one of the 10 binary mixture groups ('clusters') defined in Section 3.3. For all groups we used the same. fraction of training to test set, as for the complete database ($N_{tr,group}/N_{ts,group} = 9$). Implementing the KS algorithm individually for each group is a way to ensure that all groups are represented in the training set, even if the number of available data per group is limited [76] (e.g. binary mixtures with two non-associating, quadrupolar components). Since the KS algorithm uses the Euclidean norm of the descriptor vector, the results are sensitive to descriptor scaling [77]. For better scaling of the highly irregular multidimensional descriptor space, we used the standardized values of the descriptors [78].

$$\hat{D}_{L,m} = \frac{D_{L,m} - \overline{D}_L}{S_{DL}} \tag{24}$$

with $\hat{D}_{L,m}$ the standardized value of descriptor D_L for mixture m, $\overline{D}_L$ denotes the average value of descriptor D_L for all mixtures, and S_{DL} is the standard deviation of descriptor D_L for all mixtures of the database. In our implementation, the KS algorithm is initialized on the boundaries of the multidimensional descriptor space. Thus, the designed training set is expected to be sensitive to outliers [76].

We exclude outliers from the training set iteratively in step 3. We fit the model coefficients N_m times, excluding one mixture β at a time ("Leave-One-Out" procedure), and we calculate the corresponding sum of the residuals $r_{N_m-1,\beta}$

$$r_{N_m-1,\beta} = \sum_{\substack{m=1 \\ m\neq\beta}}^{N_m-1} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2 \tag{25}$$

with $\beta \in \{1,2,\ldots,N_m\}$. We determine the mean value

$$\overline{r}_{N_m-1} = \frac{1}{N_m} \cdot \sum_{\beta}^{N_m} r_{N_m-1,\beta} \tag{26}$$

and the standardized value of the residual for each mixture β around the mean value $\overline{r}_{N_m-1}$ as

$$\hat{r}_{N_m-1,\beta} = \frac{r_{N_m-1,\beta} - \overline{r}_{N_m-1}}{S_r} \tag{27}$$

where S_r is the standard deviation of the residuals. Mixtures that led to $\hat{r}_{N_m-1,\beta}$ values greater than three standard deviation units ($3S_r$) were considered as outliers and were excluded [70,71,79].

3.4.3. Model validation

Model validation is an integral part of any QSPR method. On the one hand, the goodness-of-fit, the robustness of the model and its internal predictive power are measured by internal validation techniques. On the other hand, the actual predictive power of the model is evaluated by external validation. External validation is defined through data that was not used for the model regression [72].

The goodness-of-fit is measured by the coefficient of multiple determination R^2. The coefficient of multiple determination estimates the portion of the variation in the independent variable (here k_{ij}^{fit}) that is explained by the regression [72,73,79]. R^2 is calculated over all mixtures of the training set (N_{tr}) and is defined as

$$R^2 = 1 - \frac{\sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{\text{QSPR}}\right)^2}{\sum_{m=1}^{N_{\text{tr}}} \left(k_{ij,m}^{\text{fit}} - \overline{k}_{ij}^{\text{fit}}\right)^2} \tag{28}$$

where $\overline{k}_{ij}^{\text{fit}}$ is the average value of k_{ij}^{fit} for all mixtures of the training set. A QSPR model with R^2 higher than 0.6 can be considered predictive [73]. However, Tropsha [73] also points out, that the predictive power of a QSPR model needs to be further evaluated for compounds that were not included in the training set.

The Leave-One-Out (LOO) cross-validated correlation coefficient Q^2_{LOO} and the Leave-Many-Out (LMO) cross-validated correlation coefficient Q^2_{LMO} measure the robustness of the model and its internal predictive power. The cross-validated correlation coefficients should be calculated over a large number of trials. In both the LOO and LMO procedures, a certain subset of the training set is omitted at each trial.

For the calculation of the LOO cross-validated correlation coefficient Q^2_{LOO} the number of trials is equal to the total number of mixtures in the QSPR database N_m. At each trial a mixture is individually omitted. The QSPR model coefficients are regressed using the remaining $N_m - 1$ mixtures as training set. Using this regression, one can now predict the k_{ij} value ($k_{ij}^{QSPR/-m}$) for mixture m (which was temporarily excluded from the training set). The definition of Q^2_{LOO} according to the guidelines of the Organization of Economic Co-operation and Development (OECD) [72] is

$$Q^2_{\text{LOO}} = 1 - \frac{\sum_{m=1}^{N_m} \left(k_{ij,m}^{\text{fit}} - k_{ij,m}^{QSPR/-m}\right)^2}{\sum_{m=1}^{N_m} \left(k_{ij,m}^{\text{fit}} - \overline{k}_{ij}^{\text{fit}}\right)^2}. \tag{29}$$

At each trial of the LMO procedure, more mixtures than one are simultaneously excluded. Here, we have calculated the value of Q^2_{LMO} over $2N_m$ number of trials. At each trial, 50% of the complete database was excluded [71] ($N_{\text{LMO}} = 0.5N_m$). The value of the LMO cross-validated correlation coefficient for each trial $Q^2_{LMO/50,trial}$ is calculated as

$$Q^2_{LMO/50,trial} = 1 - \frac{\sum_{m=1}^{N_{LMO}} \left(k_{ij,m}^{fit} - k_{ij,m}^{QSPR/-m}\right)^2}{\sum_{m=1}^{N_m} \left(k_{ij,m}^{fit} - \bar{k}_{ij}^{fit}\right)^2} \quad (30)$$

with $k_{ij,m}^{QSPR/-m}$ representing the predicted k_{ij} value for each mixture m that is part of the excluded set. The standard deviation and the average value of $Q^2_{LMO/50,trial}$ over all trials are used as indicators of the robustness in the model prediction. A robust model should remain invariant to changes of the training set and it is thus expected to exhibit small difference between the average value of Q^2_{LMO} and R^2 [80]. We should note that the LMO procedure demands to generate as many different training sets as the number of trials. The Kennard and Stone algorithm described in Section 3.4.2 is initialized on a fixed point of the descriptor space and therefore provides a single design for the training set. Thus, for the calculation of Q^2_{LMO}, we used a straightforward random selection of the training set.

An important indicator of the actual predictive capability of the model is the external explained variance Q^2_{ext}. The external explained variance is a measure for the quality of prediction for data that were not included in the set used for the model development. There are several approaches in the literature for the calculation of Q^2_{ext} [71,72,81,82]. In our implementation, the training set is much larger than the test set. In this case, the definition of Consonni et al. [81] is appropriate, with

$$Q^2_{ext} = 1 - \frac{\left[\sum_{m=1}^{N_{ts}} \left(k_{ij,m}^{fit} - k_{ij,m}^{QSPR}\right)^2\right] \Big/ N_{ts}}{\left[\sum_{m=1}^{N_{tr}} \left(k_{ij,m}^{fit} - \bar{k}_{ij}^{fit}\right)^2\right] \Big/ N_{tr}}. \quad (31)$$

The numerator in Eq. (31) sums over data of the test set, whereas the denominator sums over the training set.

4. Results

4.1. Estimation of k_{ij} based on London's dispersive theory for mixtures

Calculating k_{ij} values from London's dispersive theory, according to Eq. (5), requires segment diameters σ_i and σ_j and the ionization potentials I_i and I_j. Experimental values of the ionization potentials were taken from the literature [83]. We limit consideration to mixtures of two non-polar, non-associating components (group 1). Fig. 1 compares the values of k_{ij} calculated with Eq. (5) to the values of k_{ij} individually adjusted on experimental data. A detailed list of the examined binary mixtures with the corresponding values of $k_{ij}(\sigma,I)$ and k_{ij}^{fit} is given in Table S1 of the Supporting Information.

k_{ij} calculated with Eq. (5) cannot take on negative values. For mixtures of non-polar, non-associating components this limitation is not strongly restricting the results, because the negative k_{ij}^{fit} values are all fairly close to zero. Fig. 2 shows that for the mixtures of group 1 with negative values of k_{ij}^{fit} (mixtures #1,9,15,19,26,30,34,44,45) the results of phase equilibrium calculations remain good for calculations with the slightly positive k_{ij} values from Eq. (5). Mixtures with polar and associating components, though, often demand more negative k_{ij} values. In this case, the predictions of k_{ij} with Eq. (5) are expected to lead to more significant errors.

Higher deviations in the prediction of k_{ij} are observed for binary mixtures including carbon monoxide (mixtures 2, 3 and 4 in Figs. 1 and 2) and for binary mixtures including methane (mixtures 8, 13 and 27). These mixtures demand higher values of k_{ij} for an adequate description. We conjecture that in those cases, k_{ij} does not only correct the dispersive intermolecular potential of the mixture but compensates other model deficiencies. Below we show that a QSPR model based on semi-empirical descriptors takes the necessary contributions into account to model the value of k_{ij} for these mixtures.

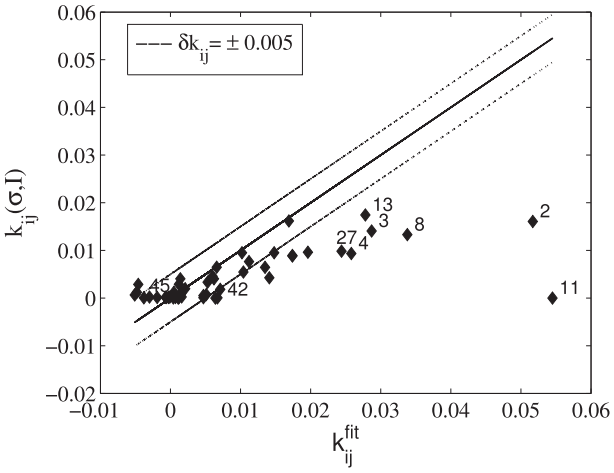

Fig. 1. Comparison of $k_{ij}(\sigma,I)$ values estimated from the London's dispersive theory (Eq. (5)) to k_{ij}^{fit} values individually adjusted to experimental data for mixtures of two non-polar, non-associating components (group 1). The dashed lines illustrate the absolute deviations of ±0.005 for the value of k_{ij}^{fit}.

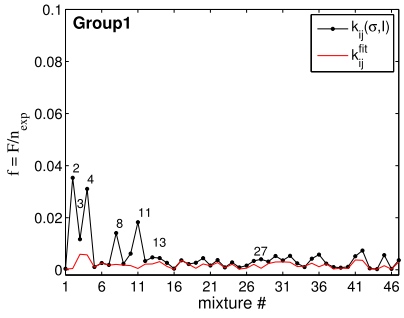

Fig. 2. Deviations of PC-SAFT EoS from experimental binary phase equilibrium data of two non-associating, non-polar components (group 1). The deviations are defined as $f = F/n_{exp}$ according to Eq. (23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The symbols (connected by black line) are obtained using estimated values $k_{ij}(\sigma,I)$. For hexadecane no ionization potential σ was available, which is why $k_{ij}(\sigma,I)$ was set to zero for mixtures #12 and #43. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

4.2. QSPR estimation of k_{ij} - model regression and assessment of the results

The descriptors of the QSPR model for the estimation of k_{ij} values were constructed using the PC-SAFT molecular parameters. The initial structure of the descriptors was defined in an ad hoc manner and their exact structure was adjusted empirically, by varying exponents, as described in Section 3.1. In the course of this process, different sets of descriptors were generated. For each set of descriptors, the coefficients of the QSPR model were regressed as described in Section 3.4.1. The performance of each QSPR model was evaluated using the internal and external validation techniques described in Section 3.4.3. The training and the test set remained unchanged. The QSPR model that achieved the best overall

description for all groups of mixtures based on the value of R^2 has been obtained with the descriptors defined in Eqs. (8)–(17). Here, we limit our discussion to the results of this model.

Mixtures with associating components are the most difficult to describe. For the regression of the model coefficients, we therefore distinguish two cases: In the first case, we estimate the model coefficients only for mixtures of non-associating components ("1st regression case": groups 1 to 6). In the second case, we adjust the model only to mixtures containing associating components ("2nd regression case": groups 7 to 10).

In the first case, the database of mixtures with only non-associating components includes 199 binary mixtures. Unreliable data and outliers have been removed according to the rationale described in Section 3.4.2. Then, the training set consists of 151 and the test set of 22 binary mixtures. The adjusted model coefficients c_L along with the values of their 95% confidence intervals $\Delta c_{L,95}$ are given in Table 2. For all regressed model coefficients we observed $|\Delta c_{L,95}| \ll |c_L|$, which qualifies a stable regression model [84]. For mixtures of groups 1 to 6 the values of k_{ij} predicted with the QSPR model are given in the Supporting Information (Tables S3–S8).

In the second case, for associating mixtures the database includes 253 mixtures with at least one associating component. After unreliable data and outliers have been removed, the training set consists of 167 and the test set of 23 binary mixtures. The adjusted model coefficients c_L for mixtures with at least one associating component and their 95% confidence intervals $\Delta c_{L,95}$ are given in Table 3. For this "2nd regression case" we also find $|\Delta c_{L,95}| \ll |c_L|$. The regression model for mixtures of the groups 7 to 10 can therefore be characterized stable as well. For mixtures of groups 7 to 10 the predicted values of k_{ij} are given in the Supporting Information (Tables S9–S12).

The multiple correlation coefficient R^2 for the QSPR model adjusted to mixtures of non-associating components is 88.4%. For the QSPR model adjusted to mixtures of associating components R^2 is 70.1%. In both cases, the model can, in the jargon of QSPR methods, be considered to be predictive. Fig. 3 shows the values of the multiple correlation coefficient (R^2_{group}) calculated individually for each group of binary mixtures. Values of R^2_{group} have been calculated using Eq. (20), with the summation running over all mixtures of the training set for each particular group. The results of the internal and external validation of the QSPR model for both regression cases are summarized in Table 4.

The QSPR model adjusted to mixtures of non-associating components describes all types of non-associating mixtures reasonably well. The results are satisfactory, especially considering the diversity of mixtures in our database. The value of the LOO cross-validated coefficient Q^2_{LOO} and the average value of the LMO cross-validated coefficient $\widehat{Q}^2_{\mathrm{LMO/50}}$ (Table 4) are close to the value of R^2. The proposed model for mixtures with non-associating, polar or non-polar components can be thus characterized as sufficiently robust [80]. Further, the low standard-deviation of $Q^2_{\mathrm{LMO/50}}$ (5.4%) over 346 trials indicates that the model is robust against strong variations of the training set.

Table 2
Results of the regressed model coefficients for mixtures of non-associating components (groups 1 to 6). The values of the coefficients c_L and their 95% confidence interval $\Delta c_{L,95}$ for the QSPR model defined in Eqs. (19) and (21), with $c^{\mathrm{assoc,s}} = 0$, $c^{\mathrm{assoc,c}} = 0$ and $c^{\mathrm{assoc,sc}} = 0$.

Mixtures with non-associating components		
	$c_L \cdot 10^4$	$\Delta c_{L,95} \cdot 10^4$
c^{LJ}	−103.285	±0.065
$c^{\mathrm{dd,a}}$	57.533	±0.235
$c^{\mathrm{dd,b}}$	−25.064	±0.095
c^{qq}	0.114	±0.001
c^{dq}	−0.645	±0.002

Table 3
Results of the regressed model coefficients for mixtures with at least one associating component (groups 7 to 10). The values of the coefficients c_L and their 95% confidence interval $\Delta c_{L,95}$ for the QSPR model defined in Eqs. (19) and (21).

Mixtures with associating components		
	$c_L \cdot 10^4$	$\Delta c_{L,95} \cdot 10^4$
c^{LJ}	−149.236	±0.591
$c^{\mathrm{dd,a}}$	20.178	±0.841
$c^{\mathrm{dd,b}}$	−23.180	±0.210
c^{qq}	−0.428	±0.006
c^{dq}	0.089	±0.006
$c^{\mathrm{assoc,s}}$	−12.432	±0.062
$c^{\mathrm{assoc,c}}$	−4.797	±0.018
$c^{\mathrm{assoc,sc}}$	0.592	±0.006

For the model adjusted to mixtures of associating components ("2nd regression case") the coefficients of multiple determination R^2_{group} (explained variance) for the individual groups are also high, particularly for the groups 7, 9 and 10. Group 8 contains mixtures of species with widely differing pure component parameters which we call asymmetry. To properly and transferably account for this asymmetry in the resulting k_{ij} values was only partially successful in our QSPR approach. We note that mixtures of group 8 represent 34% of the training set used for the adjustment of the QSPR model coefficients in the "2nd regression case". When mixtures of group 8 are omitted during the Leave-One-Out (LOO) or Leave-Many-Out (LMO) validation tests, the QSPR model demonstrates a much better overall performance in terms of the Q^2_{LOO} and $\widehat{Q}^2_{\mathrm{LMO/50}}$ QSPR validation measures. The model adjusted to mixtures of associating components is more uncertain than the QSPR model for mixtures of non-associating components. The choice of the training set acts on the result. This is reflected directly in a) the high standard deviation

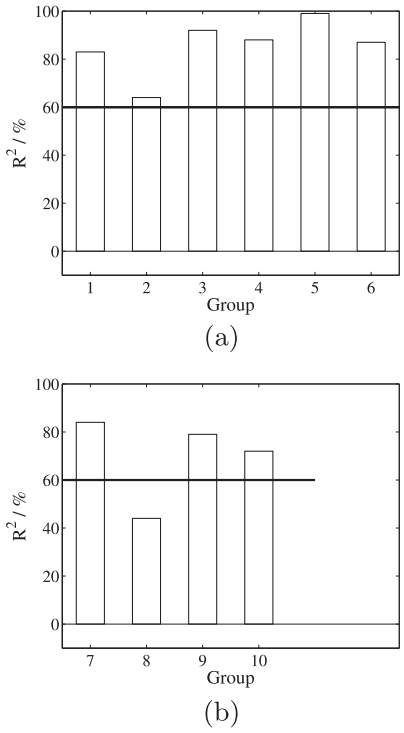

Fig. 3. Coefficient of multiple determination R^2 for: (a) mixtures of two non-associating components (groups 1 to 6) and (b) mixtures including at least one associating component (groups 7 to 10).

Table 4
Results of internal and external validation of the QSPR model (Eq. (19)) developed for mixtures of non-associating and mixtures of associating components.

Mixtures with non-associating components			
R^2(%)	Q^2_{LOO}(%)	$\widehat{Q}^2_{LMO/50}$(%)	Q^2_{ext}(%)
88.4	86.5	84.5	92.4
Mixtures with associating components			
R^2(%)	Q^2_{LOO}(%)	$\widehat{Q}^2_{LMO/50}$(%)	Q^2_{ext}(%)
70.1	64.3	57.9	82.3

of the $Q^2_{LMO/50}$ coefficient over 380 trials (22.6%) and b) the higher difference between the value of R^2 and the values of the cross validated coefficients $\widehat{Q}^2_{LMO/50}$ and Q^2_{LOO} (Table 4).

The external explained variance Q^2_{ext} by the QSPR model adjusted to mixtures of non-associating components is 92%. For the QSPR model adjusted to mixtures of associating components Q^2_{ext} is 82%. The high values of Q^2_{ext} indicate the high predictive power of the QSPR model in both cases. Nevertheless, the results of Q^2_{ext} should be interpreted cautiously. One should bear in mind that the size of the test set is much smaller and less diverse than the size of the training set ($N_{tr}/N_{ts} = 9$).

Generally, the proposed QSPR model is stable and robust and has (in the sense of a QSPR method) a good predictive capacity. The statistical measures (Table 4) indicate that the QSPR model predicts k_{ij} values in binary mixtures of non-associating components with sufficient accuracy. For mixtures of associating components, the model is also reliable (in the jargon of QSPR). However, the estimations of k_{ij} for mixtures containing one dipolar, non-associating component and one associating component are more uncertain and should be used with reservation.

4.3. Evaluating phase equilibria with predicted k_{ij} values

The k_{ij} values used for the regression of the QSPR model (k_{ij}^{fit}) are not experimental data, but parameters fitted to experimental data. We therefore find it important to evaluate the QSPR model beyond just the statistical measures of the QSPR method. We have examined the results of phase equilibrium calculations with the predicted k_{ij} values for each group of mixtures. The parity plots and the diagrams that show the results in phase equilibrium calculations for all groups of mixtures are given in the Supporting Information (Figs. S1–S4).

Here, we discuss in more detail the results for mixtures of group 4 (mixtures of one non-associating, quadrupolar and one non-associating, dipolar component). We find the performance for mixtures of group 4 representative of the overall performance of the QSPR predictions of k_{ij} for non-associating mixtures ($R^2_{group\ 4} = 88.1\% \simeq R^2$). Moreover, the prediction for mixtures of group 4 is of practical interest, for example by the application of CAMD for polar solvents, for the CO_2 capture with physical absorption [18].

Fig. 4 compares the estimated values, k_{ij}^{QSPR}, to the reference values, k_{ij}^{fit}, for group 4. Further, Fig. 5 illustrates the accuracy in phase equilibrium calculations with PC-SAFT for the mixtures of group 4 with k_{ij} values estimated with the QSPR model adjusted to non-associating mixtures (Table 2). The results in phase equilibrium calculations are compared to experimental VLE data. We consider the same experimental data that were used for the individual fitting of k_{ij} values. We assess the results using the average value f of the combined residuals in the liquid mole fraction and the pressure, over the number of experimental data points for each mixture as defined in Eq. (23). In Fig. 5, we also present the results in phase equilibrium calculations, when no correction is used for the PC-SAFT EoS ($k_{ij} = 0$), as well as the results achieved when k_{ij} is individually adjusted to the experimental data, namely $f(k_{ij}^{fit})$. We observe that for the majority of the mixtures in group 4 (in both the training and the test set) phase equilibrium calculations with the QSPR estimations of k_{ij} give lower residuals than if no correction is used ($k_{ij} = 0$).

It is however difficult to quantify the relationship between the error in Ref. k_{ij} prediction $\Delta k_{ij} = \left|k_{ij}^{fit} - k_{ij}^{QSPR}\right|$ and the subsequent error in phase equilibrium calculations $\Delta f = \left|f(k_{ij}^{fit}) - f(k_{ij}^{QSPR})\right|$. For different mixtures, the sensitivity of the objective function F (and thereby f) to the value of k_{ij} is different, depending on different factors, like azeotropy or supercritical regions. Let us, for example, examine the performance of the k_{ij} prediction and phase equilibrium calculations for mixture #4 (carbon dioxide - 1-bromobenzene) compared to mixture #20 (1-hexene - ethylene). For mixture #4, the error in the estimation of k_{ij} is much higher than the error in the estimation of k_{ij} for mixture #20 (Fig. 4, Table S5). Still, the error in phase equilibrium calculations is almost identical for both mixtures (Fig. 5).

Finally, the k_{ij} values estimated from London's dispersive theory $k_{ij}(\sigma,I)$ (Eq. (5)) were used in phase equilibrium calculations for the mixtures of group 4. A list of the estimated values $k_{ij}(\sigma,I)$ is given in Table S2 in the Supporting Information. In Fig. 6, phase equilibrium calculations carried out with the estimated values k_{ij}^{QSPR} are compared to phase equilibrium calculations with the estimated values from London's dispersive theory $k_{ij}(\sigma,I)$. The results obtained with the estimated value k_{ij}^{QSPR} are in most cases better than the results obtained with $k_{ij}(\sigma,I)$. For example, mixture #7 (dimethyl ether-carbon dioxide) requires a negative k_{ij} value. Negative k_{ij} values cannot be calculated using Eq. (5). As discussed in Section 4.1, this limitation has a higher impact on mixtures containing polar components. For mixture #7, the estimated k_{ij}^{QSPR} is negative. With k_{ij}^{QSPR}, the PC-SAFT EoS is corrected and phase equilibrium calculations are considerably better. Moreover, for mixtures that require high positive k_{ij} values, the predicted value k_{ij}^{QSPR} leads to significantly lower residuals in the phase equilibrium calculations. Representative examples are mixture #2 (nitrogen-propylbenzene) and mixture #17 (nitrogen-toluene).

The k_{ij} values predicted by the proposed QSPR model (with the descriptors defined in Eqs. (8)–(17) and the regressed coefficients

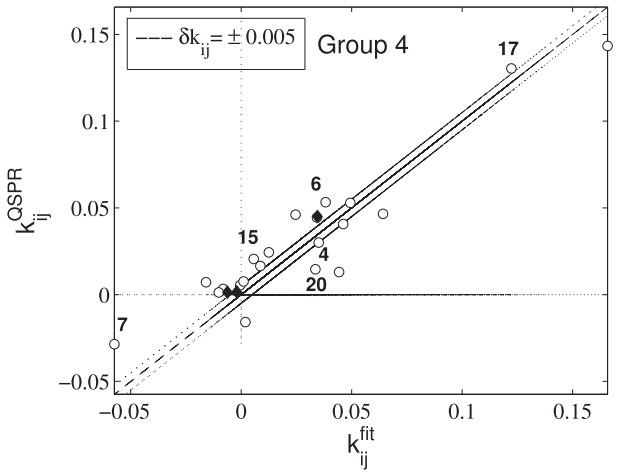

Fig. 4. Comparison of the estimated k_{ij}^{QSPR} values with k_{ij}^{fit} values individually adjusted to experimental data for mixtures of one non-associating, dipolar and one non-associating, quadrupolar component (group 4). Open symbols indicate mixtures of the training set, whereas solid symbols correspond to the mixtures of the test set. The dashed lines illustrate the absolute deviations of ±0.005 for the value of (k_{ij}^{fit}).

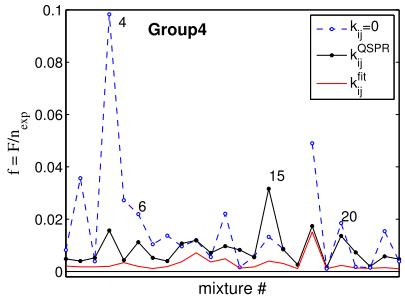

Fig. 5. Deviations of PC-SAFT from experimental binary phase equilibrium data of one non-associating, dipolar and one non-associating, quadrupolar component (group 4). The deviations are defined as $f = F/n_{exp}$ according to Eq. (23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The symbols (connected by black line) are obtained using estimated values k_{ij}^{QSPR}. The blue dashed line represent phase equilibrium calculations, when the PC-SAFT EoS is not corrected ($k_{ij} = 0$). (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

given in Tables 2 and 3) are used in phase equilibrium calculations for all mixtures considered in our QSPR database. In Fig. 7, a general overview is given of the improvement potential in phase equilibrium calculations, using the k_{ij} values estimated with the proposed QSPR model. For each group of mixtures, Fig. 7 displays the percentage of cases in which phase equilibrium calculations are more accurate with the estimated value k_{ij}^{QSPR}. The percentage of cases when phase equilibrium calculations with $k_{ij} = 0$ are more accurate and the percentage of cases when phase equilibrium calculations are equally good for k_{ij}^{QSPR} and $k_{ij} = 0$ are given as well. For all groups of mixtures (associating and non-associating) the proposed QSPR model provides k_{ij} estimates that generally improve the accuracy of PC-SAFT in phase equilibrium calculations. According to Fig. 7, the most significant improvement is achieved for mixtures of group 5. However, the small size of group 5 (Table 1) and its' limited diversity (Table S7) do not allow us to draw a generalised conclusion about the performance of the method for mixtures of two non-associating, quadrupolar components. Further, in the group with the least improvement, for mixtures of two dipolar, non-associating components (group 6), the individually adjusted k_{ij} values are close to zero (see Supporting Information). Calculation results for $k_{ij} = 0$ are thus close to optimal, so that the use of estimated k_{ij}^{QSPR} cannot lead to significant improvements for group 6.

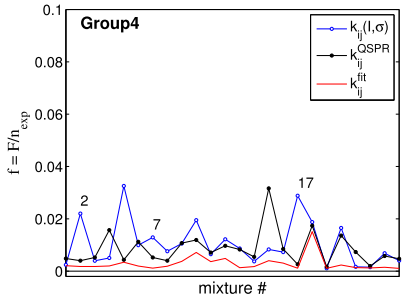

Fig. 6. Deviations of PC-SAFT from experimental binary phase equilibrium data of one non-associating, dipolar and one non-associating, quadrupolar component (group 4). The deviations are defined as $f = F/n_{exp}$ according to Eq. (23). The red line serves as a reference and is obtained for individually optimized k_{ij}^{fit} values. The black line represents phase equilibrium calculations, when the PC-SAFT EoS is corrected with k_{ij}^{QSPR}. The blue solid line is for results using $k_{ij}(\sigma,I)$ from London's dispersive theory. (\pretation of the references to color in this figure caption, the reader is referred to the web version of this article.)

5. Conclusions

We developed and analyzed a multivariate regression model for the estimation of the binary interaction parameter k_{ij} of the PC-SAFT EoS using a QSPR method. The descriptors of the regression model are based on the pure component parameters of the PC-SAFT model. The regression models were developed separately for mixtures with non-associating components and for mixtures with at least one associating component.

For mixtures with non-associating components, values of k_{ij} were also estimated from London's dispersive theory, using the expression of Hudson and McCoubrey [50]. Values of k_{ij} predicted with the QSPR model (k_{ij}^{QSPR}) and values of k_{ij} estimated from London's dispersive theory ($k_{ij}(\sigma,I)$) were compared to values of k_{ij} individually fitted on experimental data. For mixtures containing non-associating, polar components the QSPR regression model allows for negative values of k_{ij} and leads to more accurate predictions than the expression of Hudson and McCoubrey. The QSPR regression model considers additional contributions to the value of k_{ij} than just the asymmetry in the dispersive intermolecular forces.

The coefficient of multiple determination R^2 of the QSPR regression model for mixtures with non-associating components is 88.4%. For mixtures with at least one associating component R^2 is 70.1%. Thus, for both regression cases the QSPR model can, in the jargon of QSPR methods, be characterized as stable and robust and with good predictive capacity.

The k_{ij} values estimated as function of the pure component PC-SAFT parameters, with the proposed QSPR regression model, can be used to enhance the accuracy of phase equilibrium calculations with PC-SAFT. With the proposed QSPR model for the estimation of k_{ij}, no prior knowledge of the real behavior of the examined binary mixture is demanded. The QSPR approach can be implemented in a CAMD framework that uses the pure component parameters of the

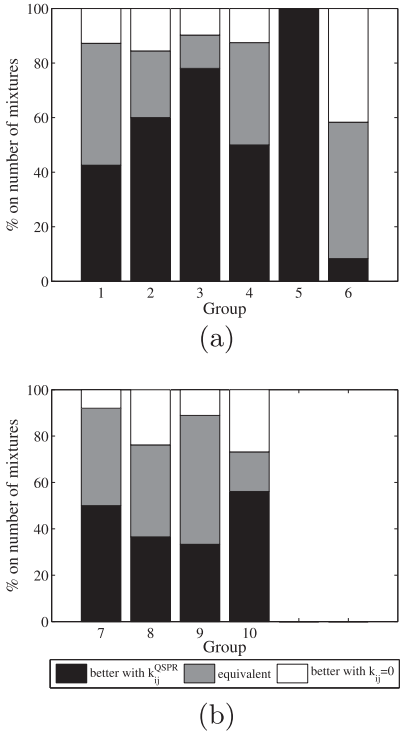

Fig. 7. Assessing phase equilibrium calculations: Black surface represents the percentage of cases for which calculations with estimated k_{ij}^{QSPR} values are more accurate than with $k_{ij} = 0$. The gray shaded area represents the percentage of cases when both cases are equivalent.

optimized fluid directly as optimization variables. In that case, we expect that phase equilibrium calculations and thus the accuracy of CAMD results can be enhanced.

Acknowledgment

This study has been performed within the CO_2 Catch-up R&D program aimed at demonstrating and optimizing pre-combustion CO_2 capture technology for the energy sector. This program is executed in a consortium of Nuon (part of Vattenfall), TU Delft and ECN.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.fluid.2015.12.016.

Nomenclature

Abbreviations

SAFT	Statistical Associating Fluid Theory
PC-SAFT	Perturbed Chain − Statistical Associating Fluid Theory
EoS	Equation of State
QM	Quantum mechanical
CAMD	Computer Aided Molecular Design
CoMT	Continuous Molecular Targeting
CAMPD	Computer Aided Molecular and Process

Design

GC	Group contribution
ORC	Organic Rankine cycle
QSPR	Quantitative Structure Property Relationship
nAnP	Non-associating, non-polar
nAdP	Non-associating, dipolar
nAqP	Non-associating, quadrupolar
AdP	Associating, dipolar
VLE	Vapor–liquid equilibrium
DDB	Dortmund Database
KS	Kennard and Stone algorithm
LOO	Leave-One-Out
LMO	Leave-Many-Out
OECD	Organization of Economic Co-operation and Development

References

[1] E.A. Muller, K.E. Gubbins, Molecular-based equations of state for associating fluids: a review of SAFT and related approaches, Ind. Eng. Chem. Res. 40 (10) (2001) 2193–2211, http://dx.doi.org/10.1021/ie000773w.
[2] I.G. Economou, Statistical associating fluid theory: a successful model for the calculation of thermodynamic and phase equilibrium properties of complex fluid mixtures, Ind. Eng. Chem. Res. 41 (5) (2002) 953–962, http://dx.doi.org/10.1021/ie0102201.
[3] S.P. Tan, H. Adidharma, M. Radosz, Recent advances and applications of statistical associating fluid theory, Ind. Eng. Chem. Res. 47 (21) (2008) 8063–8082, http://dx.doi.org/10.1021/ie8008764.
[4] C. McCabe, A. Galindo, Chapter 8 SAFT associating fluids and fluid mixtures, in: Applied Thermodynamics of Fluids, R. Soc. Chem. (2010) 215–279, http://dx.doi.org/10.1039/9781849730983-00215.
[5] W.G. Chapman, S.G. Sauer, D. Ting, A. Ghosh, Phase behavior applications of SAFT based equations of state: from associating fluids to polydisperse, polar copolymers, Fluid Phase Equilib. 217 (2004) 137–143, http://dx.doi.org/10.1016/j.fluid.2003.05.001.
[6] G. Sadowski, Thermodynamics of polymer systems, Macromol. Symp. 206 (1) (2004) 333–346, http://dx.doi.org/10.1002/masy.200450226.
[7] F. Tumakaka, J. Gross, G. Sadowski, Thermodynamic modeling of complex systems using PC-SAFT, Fluid Phase Equilib. 228–229 (0) (2005) 89–98, http://dx.doi.org/10.1016/j.fluid.2004.09.037. PPEPPD 2004 Proceedings.
[8] B. Liebergesell, S. Kaminski, C. Pauls, T.W. de Loos, T. Vlugt, K. Leonhard, A. Bardow, High-pressure vapor-liquid equilibria of the second generation biofuel blends (2-methylfuran+iso-octane) and (2-methyltetrahydrofuran+di-n-butyl ether): Experiments and PCP-SAFT modeling, Fluid Phase Equilib. 400 (0) (2015) 95–102, http://dx.doi.org/10.1016/j.fluid.2015.05.002.
[9] R. Gani, J.P. O'Connell, Properties and CAPE: from present uses to future challenges, Comput. Chem. Eng. 25 (1) (2001) 3–14 doi: S0098-1354(00)00628-1.
[10] J.P. O'Connell, R. Gani, P.M. Mathias, G. Maurer, J.D. Olson, P.A. Crafts, Thermodynamic property modeling for chemical process and product engineering: some perspectives, Ind. Eng. Chem. Res. 48 (10) (2009) 4619–4637, http://dx.doi.org/10.1021/ie801535a.
[11] R. Gani, B. Nielsen, A. Fredenslund, A group contribution approach to computer-aided molecular design, AIChE J. 37 (9) (1991) 1318–1332, http://dx.doi.org/10.1002/aic.690370905.
[12] L.-Y. Ng, F.-K. Chong, N.G. Chemmangattuvalappil, Challenges and opportunities in computer-aided molecular design, Comput. Chem. Eng. 81 (2015) 115–129, http://dx.doi.org/10.1016/j.compchemeng.2015.03.009 special Issue: Selected papers from the 8th International Symposium on the Foundations of Computer-Aided Process Design (FOCAPD 2014), July 13-17, 2014, Cle Elum, Washington,USA.
[13] F.E. Pereira, E. Keskes, A. Galindo, G. Jackson, C.S. Adjiman, Integrated Design of CO_2 Capture Processes from Natural Gas, Wiley-VCH Verlag GmbH Co. KGaA, 2008, pp. 231–248, http://dx.doi.org/10.1002/9783527631292.ch8.
[14] A. Bardow, K. Steur, J. Gross, Continuous-molecular targeting for integrated solvent and process design, Ind. Eng. Chem. Res. 49 (2010) 2834–2840, http://dx.doi.org/10.1021/ie901281w.
[15] F. Pereira, E. Keskes, A. Galindo, G. Jackson, C. Adjiman, Integrated solvent and process design using a SAFT-VR thermodynamic description: high-pressure separation of carbon dioxide and methane, Comput. Chem. Eng. 35 (3) (2011) 474–491, http://dx.doi.org/10.1016/j.compchemeng.2010.06.016.
[16] M. Lampe, M. Stavrou, H.M. Bucker, J. Gross, A. Bardow, Simultaneous optimization of working fluid and process for organic rankine cycles using PC-SAFT, Ind. Eng. Chem. Res. 53 (21) (2014) 8821–8830, http://dx.doi.org/10.1021/ie5006542.
[17] A. Papadopoulos, S. Bard, A. Chremos, E. Forte, T. Zarogiannis, P. Seferlis, S. Papadokonstantakis, C. Adjiman, A. Galindo, G. Jackson, Efficient screening and selection of post-combustion CO_2 capture solvents, Chem. Eng. Trans. 39 (2014) 211–216, http://dx.doi.org/10.3303/CET1439036.
[18] M. Stavrou, M. Lampe, A. Bardow, J. Gross, Continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) for simultaneous process and solvent design for CO_2 capture, Ind. Eng. Chem. Res. 53 (46) (2014) 18029–18041, http://dx.doi.org/10.1021/ie502924h.
[19] M. Lampe, M. Stavrou, J. Schilling, E. Sauer, J. Gross, A. Bardow, Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT, Comput. Chem. Eng. 81 (2015) 278–287.
[20] C.S. Adjiman, A. Galindo, G. Jackson, Molecules matter: the expanding envelope of process design, in: J.D.S. Mario, R. Eden, G.P. Towler (Eds.), Proceedings of the 8th International Conference on Foundations of Computer-aided Process Design, Vol. 34 of Computer Aided Chemical Engineering, Elsevier, 2014, pp. 55–64, http://dx.doi.org/10.1016/B978-0-444-63433-7.50007-9.
[21] R. Gani, Chapter 14 case studies in chemical product design - use of CAMD techniques, in: R.G. Ka M. Ng, K. Dam-Johansen (Eds.), Chemical Product Design: toward a Perspective through Case Studies, Vol. 23 of Computer Aided Chemical Engineering, Elsevier, 2007, pp. 435–458, http://dx.doi.org/10.1016/S1570-7946(07)80017-4.
[22] A.I. Papadopoulos, P. Linke, Integrated solvent and process selection for separation and reactive separation systems, Chem. Eng. Process 48 (5) (2009) 1047–1060, http://dx.doi.org/10.1016/j.cep.2009.02.004.
[23] M. Harini, J. Adhikari, K.Y. Rani, A review on property estimation methods and computational schemes for rational solvent design: a focus on pharmaceuticals, Ind. Eng. Chem. Res. 52 (21) (2013) 6869–6893, http://dx.doi.org/10.1021/ie301329y.
[24] A. Gil-Villegas, A. Galindo, P.J. Whitehead, S.J. Mills, G. Jackson, A.N. Burgess, Statistical associating fluid theory for chain molecules with attractive potentials of variable range, J. Chem. Phys. 106 (1997) 4168–4186, http://dx.doi.org/10.1063/1.473101.
[25] J. Burger, V. Papaioannou, S. Gopinath, G. Jackson, A. Galindo, C.S. Adjiman, A hierarchical method to integrated solvent and process design of physical CO_2 absorption using the SAFT-γ Mie approach, AIChE J. 61 (10) (2015) 3249–3269, http://dx.doi.org/10.1002/aic.14838.
[26] V. Papaioannou, T. Lafitte, C. Avendano, C.S. Adjiman, G. Jackson, E.A. Muller, A. Galindo, Group contribution methodology based on the statistical associating fluid theory for heteronuclear molecules formed from Mie segments, J. Chem. Phys. 140 (5) (2014), http://dx.doi.org/10.1063/1.4851455.
[27] J. Gross, G. Sadowski, Perturbed-Chain SAFT: an equation of state based on a perturbation theory for chain molecules, Ind. Eng. Chem. Res. 40 (2001) 1244–1260, http://dx.doi.org/10.1021/ie0003887.
[28] J. Gross, G. Sadowski, Application of the Perturbed-chain SAFT equation of state to associating systems, Ind. Eng. Chem. Res. 41 (2002) 5510–5515, http://dx.doi.org/10.1021/ie010954d.
[29] J. Gross, An equation-of-state contribution for polar components: quadrupolar molecules, AIChE J. 51 (2005) 2556–2568, http://dx.doi.org/10.1002/aic.10502.
[30] J. Gross, J. Vrabec, An equation-of-state contribution for polar components: dipolar molecules, AIChE J. 52 (2006) 1194–1204, http://dx.doi.org/10.1002/

aic.10683.
[31] P.T. Sikora, Combining rules for spherically symmetric intermolecular potentials, J. Phys. B At., Mol. Opt. Phys. 3 (11) (1970) 1475.
[32] M.J. Hiza, A.G. Duncan, A correlation for the prediction of interaction energy parameters for mixtures of small molecules, AIChE J. 16 (5) (1970) 733–738, http://dx.doi.org/10.1002/aic.690160509.
[33] C.L. Kong, Combining rules for intermolecular potential parameters. ii. Rules for the Lennard-Jones (12-6) potential and the Morse potential, J. Chem. Phys. 59 (5) (1973) 2464–2467, http://dx.doi.org/10.1063/1.1680358.
[34] F. Kohler, J. Fischer, E. Wilhelm, Intermolecular force parameters for unlike pairs, J. Mol. Struct. 84 (34) (1982) 245–250, http://dx.doi.org/10.1016/0022-2860(82http://dx.doi.org/10.1016/0022-2860(82)85257-5.
[35] K. Tang, J. Toennies, New combining rules for well parameters and shapes of the van der Waals potential of mixed rare gas systems, Z Phys. D. At. Mol. Cl. 1(1) (1986) 91–101, http://dx.doi.org/10.1007/BF01384663.
[36] T.A. Halgren, The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters, J. Am. Chem. Soc. 114 (20) (1992) 7827–7843, http://dx.doi.org/10.1021/ja00046a032.
[37] M. Waldman, A. Hagler, New combining rules for rare gas van der Waals parameters, J. Comput. Chem. 14 (9) (1993) 1077–1084, http://dx.doi.org/10.1002/jcc.540140909.
[38] T. Schnabel, J. Vrabec, H. Hasse, Unlike Lennard-Jones parameters for vapor-liquid equilibria, J. Mol. Liq. 135 (1–3) (2007) 170–178, http://dx.doi.org/10.1016/j.molliq.2006.12.024.
[39] A.J. Haslam, A. Galindo, G. Jackson, Prediction of binary intermolecular potential parameters for use in modelling fluid mixtures, Fluid Phase Equilib. 266 (2008) 105–128, http://dx.doi.org/10.1016/j.fluid.2008.02.004.
[40] M. Singh, K. Leonhard, K. Lucas, Making equation of state models predictive: part 1: quantum chemical computation of molecular properties, Fluid Phase Equilib. 258 (2007) 16–28, http://dx.doi.org/10.1016/j.fluid.2007.05.021.
[41] K. Leonhard, N.V. Nhu, K. Lucas, Making equation of state models predictive: part 2: an improved PCP-SAFT equation of state, Fluid Phase Equilib. 258 (2007) 41–50, http://dx.doi.org/10.1016/j.fluid.2007.05.019.
[42] F.T. Peters, F.S. Laube, G. Sadowski, PC-SAFT based group contribution method for binary interaction parameters of polymer/solvent systems, Fluid Phase Equilib. 358 (2013) 137–150, http://dx.doi.org/10.1016/j.fluid.2013.05.033.
[43] D. Nguyen-Huynh, J.-P. Passarello, P. Tobaly, J.-C. de Hemptinne, Modeling phase equilibria of asymmetric mixtures using a group-contribution SAFT (GC-SAFT) with a k_{ij} 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 (2008) 8847–8858, http://dx.doi.org/10.1021/ie071643r.
[44] D. Nguyen-Huynh, T.K.S. Tran, S. Tamouza, J.-P. Passarello, P. Tobaly, J.-C. de Hemptinne, Modeling phase equilibria of asymmetric mixtures using a group-contribution SAFT (GC-SAFT) with a k_{ij} correlation method based on London's theory. 2. application to binary mixtures containing aromatic hydrocarbons, n-alkanes, CO_2, N_2, and H_2 S, Ind. Eng. Chem. Res. 47 (2008) 8859–8868, http://dx.doi.org/10.1021/ie071644j.
[45] S. Tamouza, J.-P. Passarello, P. Tobaly, J.-C. de Hemptinne, Group contribution method with SAFT EOS applied to vapor liquid equilibria of various hydrocarbon series, Fluid Phase Equilib. 222–223 (2004) 67–76, http://dx.doi.org/10.1016/j.fluid.2004.06.038. Proceedings of the Fifteenth Symposium on Thermophysical Properites.
[46] J.A.P. Coutinho, P.M. Vlamos, G.M. Kontogeorgis, General form of the cross-energy parameter of equations of state, Ind. Eng. Chem. Res. 39 (2000) 3076–3082, http://dx.doi.org/10.1021/ie990904x.
[47] G.M. Kontogeorgis, E.C. Voutsas, I.V. Yakoumis, D.P. Tassios, An equation of state for associating fluids, Ind. Eng. Chem. Res. 35 (1996) 4310–4318, http://dx.doi.org/10.1021/ie9600203.
[48] M. Shacham, G.S. Cholakov, R.P. Stateva, N. Brauner, Quantitative structure-property relationships for prediction of phase equilibrium related properties, Ind. Eng. Chem. Res. 49 (2010) 900–912, http://dx.doi.org/10.1021/ie900807j.
[49] A.M. Abudour, S.A. Mohammad, Robert 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.
[50] G.H. Hudson, J. McCoubrey, Intermolecular forces between unlike molecules, Trans. Faraday Soc. 56 (1959) 761–766.
[51] G.M. Kontogeorgis, Association theories for complex thermodynamics, Chem. Eng. Res. Des. 91 (2013) 1840–1858, http://dx.doi.org/10.1016/j.cherd.2013.07.006 the 60th Anniversary of the European Federation of Chemical Engineering (EFCE).
[52] A. Katritzky, M. Karelson, R. Petrukhin, CODESSA PRO Project, 2005. http://www.codessa-pro.com/.
[53] TALETE-SRL, Dragon 6, 2013. http://www.talete.mi.it/products/.
[54] A. Kazakov, C.D. Muzny, V. Diky, R.D. Chirico, M. Frenkel, Predictive correlations based on large experimental datasets: critical constants for pure compounds, Fluid Phase Equilib. 298 (2010) 131–142, http://dx.doi.org/10.1016/j.fluid.2010.07.014.
[55] M.A. Sobati, D. Abooali, Molecular based models for estimation of critical properties of pure refrigerants: quantitative structure property relationship (QSPR) approach, Thermochim. Acta 602 (2015) 53–62, http://dx.doi.org/10.1016/j.tca.2015.01.006.
[56] E.N. Muratov, E.V. Varlamova, A.G. Artemenko, P.G. Polishchuk, V.E. Kuzmin, Existing and developing approaches for QSAR analysis of mixtures, Mol. Inf. 31 (3–4) (2012) 202–221, http://dx.doi.org/10.1002/minf.201100129.
[57] M. Kleiner, G. Sadowski, Modeling of polar systems using PCP-SAFT: an approach to account for induced-association interactions, J. Phys. Chem. C 111 (43) (2007) 15544–15553, http://dx.doi.org/10.1021/jp072640v.
[58] G.M. Kontogeorgis, G.K. Folas, Thermodynamic Models for Industrial Applications, John Wiley & Sons, Ltd, 2009, http://dx.doi.org/10.1002/9780470747537.fmatter.
[59] J.P. Wolbach, S.I. Sandler, Using molecular orbital calculations to describe the phase behavior of cross-associating mixtures, Ind. Eng. Chem. Res. 37 (1998) 2917–2928, http://dx.doi.org/10.1021/ie970781l.
[60] N. von Solms, M.L. Michelsen, G.M. Kontogeorgis, Applying association theories to polar fluids, Ind. Eng. Chem. Res. 43 (2004) 1803–1806, http://dx.doi.org/10.1021/ie034243m.
[61] M. Kleiner, F. Tumakaka, G. Sadowski, Thermodynamic modeling of complex systems, Struct Bond. Springer Berlin Heidelberg (2008) 1–34.
[62] J.O. Hirschfelder, C.F. Curtiss, R.B. Bird, Molecular Theory of Gases, John Wiley & Sons, 1987.
[63] S.H. Huang, M. Radosz, Equation of state for small, large, polydisperse, and associating molecules, Ind. Eng. Chem. Res. 29 (11) (1990) 2284–2294, http://dx.doi.org/10.1021/ie00107a014.
[64] F. Ruether, G. Sadowski, Modeling the solubility of pharmaceuticals in pure solvents and solvent mixtures for drug process design, J. Pharm. Sci. 98 (2009) 4205–4215, http://dx.doi.org/10.1002/jps.21725.
[65] AIChE-DECHEMA, Pure, Component Database DIPPR-801, Tech. rep., AIChE-DECHEMA, 2011.
[66] D.D.B.S.S.T. GmbH, Dortmund Data Bank Software, Tech. Rep, 2015.
[67] J. Gmehling, B. Kolbe, M. Kleiber, J. Rarey, Chemical Thermodynamics for Process Simulation, Wiley-VCH Verlag Co. KGaA, 2012.
[68] G. Soave, S. Gamba, L.A. Pellegrini, SRK equation of state: predicting binary interaction parameters of hydrocarbons and related compounds, Fluid Phase Equilib. 299 (2010) 285–293, http://dx.doi.org/10.1016/j.fluid.2010.09.012.
[69] Mathworks, Matlab(r), Mathworks – Documentation Center, 05 2015. http://de.mathworks.com/help/stats/nlinfit.html.
[70] D.C. Montgomery, E.A. Peck, G.G. Vining, Introduction to Linear Regression Analysis, fifth ed., John Wiley & Sons, 2012.
[71] P. Gramatica, P. Pilutti, E. Papa, Validated QSAR prediction of OH tropospheric degradation of VOCs: splitting into training-test sets and consensus modeling, J. Chem. Inf. Comput. Sci. 44 (2004) 1794–1802, http://dx.doi.org/10.1021/ci049923u pMID: 15446838.
[72] Guidance Document on the Validation of Quantitative Structure-activity Relationships QSAR Models, 2007.
[73] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, Mol. Inf. 29 (2010) 476–488, http://dx.doi.org/10.1002/minf.201000061.
[74] M. Daszykowski, B. Walczak, D. Massart, Representative subset selection, Anal. Chim. Acta 468 (2002) 91–103, http://dx.doi.org/10.1016/S0003-2670(02)00651-7.
[75] R.W. Kennard, L. Stone, Computer aided design of experiments, Technometrics 11 (1969) 137–148.
[76] P. de Groot, G. Postma, W. Melssen, L. Buydens, Selecting a representative training set for the classification of demolition waste using remote NIR sensing, Anal. Chim. Acta 392 (1999) 67–75, http://dx.doi.org/10.1016/S0003-2670(99)00193-2.
[77] B. Bourguignon, P.F. de Agular, M.S. Khots, D.L. Massart, Optimization in irregularly shaped regions: pH and solvent strength in reversed-phase high-performance liquid chromatography separations, Anal. Chem. 66 (1994) 893–904.
[78] B.G. Tabachnick, L.S. Fidell, Using Multivariate Statistics, Pearson, 2013.
[79] A.R. Katritzky, M. Kuanar, S. Slavov, C.D. Hall, M. Karelson, I. Kahn, D.A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction, Chem. Rev. 110 (2010) 5714–5789, http://dx.doi.org/10.1021/cr900238d pMID: 20731377.
[80] V. Prana, P. Rotureau, G. Fayet, D. Andre, S. Hub, P. Vicot, L. Rao, C. Adamo, Prediction of the thermal decomposition of organic peroxides by validated QSPR models, J. Hazard. Mater. 276 (2014) 216–224, http://dx.doi.org/10.1016/j.jhazmat.2014.05.009.
[81] V. Consonni, D. Ballabio, R. Todeschini, Evaluation of model predictive ability by external validation techniques, J. Chemom. 24 (2010) 194–201, http://dx.doi.org/10.1002/cem.1290.
[82] A. Golbraikh, A. Tropsha, Beware of q2!, J. Mol. Graph. Modell. 20 (2002) 269–276, http://dx.doi.org/10.1016/S1093-3263(01)00123-1.
[83] CRC handbook of chemistry and physics, Internet Version 2005, 2005. http://www.hbcpnetbase.com/.
[84] M. Hechinger, Model-based Identification of Promising Biofuel Candidates for Spark-ignited Engines, Ph.D. thesis, Rheinisch-Westfalischen Technischen Hochschule Aachen, 2014.