Available online at www.sciencedirect.com
SCIENCE DIRECT
Fluid Phase Equilibria 246 (2006) 39–51
FLUID PHASE EQUILIBRIA
www.elsevier.com/locate/fluid

SVRC–QSPR model for predicting saturated vapor pressures of pure fluids

Srinivasa S. Godavarthy, Robert L. Robinson Jr., Khaled A.M. Gasem *

School of Chemical Engineering, Oklahoma State University, Stillwater, OK 74078, United States

Received 13 February 2006; received in revised form 22 May 2006; accepted 23 May 2006
Available online 2 June 2006

Abstract

Knowledge of thermo-physical properties of organic chemicals is essential to chemical and process design applications. Vapor pressure is one such property used directly in process calculations and as input to property-prediction models. Although experimental determination of vapor pressures remains an option, often it is not possible to measure vapor pressure data experimentally for toxic or yet to be synthesized molecules. Current vapor pressure models, which utilize traditional physical properties as inputs, are limited by their range of applicability and/or by poor suitability for generalization. Further, recent quantitative structure–property relations (QSPR) models for vapor pressure have been limited to single-temperature generalizations (e.g., 298 K); thus, the distinct advantages offered by advances in computational chemistry as they relate to structure–property model generalizations have not been fully realized.

In this study, we present an integrated approach for developing a generalized model which is capable of predicting accurately the vapor pressure of organic chemicals over the entire saturation range (the triple point to the critical point). The approach uses a theoretical framework to develop the fluid behavior model and QSPR to generalize the parameters of the model. Specifically, we first apply our scaled variable reduced coordinates (SVRC) model to a diverse dataset containing over 1221 molecules involving 73 classes of chemicals, and then we generalize the SVRC parameters using structure–property (SP) models. For this modeling effort, reliable experimental vapor pressure data were obtained from the DIPPR database. The results for 52,445 data points indicate that: (a) the SVRC model represents these saturated vapor pressure data with 0.35% average absolute deviation (AAD), and (b) the generalized SVRC–QSPR model predicts the saturated vapor pressures with 0.5% AAD.

Keywords: Vapor pressure; Pure fluid; Models; QSPR; SVRC

1. Introduction

Vapor pressure is an important thermo-physical property in numerous chemical process and product design applications [1]. Knowledge of pure-fluid vapor pressure is also essential to understanding fluid phase behavior and to performing multiphase equilibrium calculations of multi-component systems [2].

Traditionally, vapor pressure has been determined experimentally by techniques of varying complexities, and experimental measurements are still needed when dealing with challenging fluids. However, experimental determination of vapor pressures of the ever increasing number of chemicals requires significant time and cost investments. These constraints are further amplified in the case of hazardous chemicals [3].

Many correlations for estimating vapor pressures have been used to complement existing experimental measurements [4]. However, most of these correlations suffer from a limited range of applicability and/or poor suitability for generalization. The majority of the existing vapor pressure correlations are based on the corresponding states theory (CST) [5–7], either as direct pressure–temperature ($p(T)$) CST correlations or as an element of equations-of-state (EOS) [8–10] formulations. Although the EOS approach is generally more efficient in correlating the thermodynamic behavior of normal fluids, most current EOS are incapable of predicting accurately the saturation properties of polar and heavy hydrocarbons. Further, they often exhibit large prediction errors near the critical point and the triple point of the fluid. As a result of these inadequacies, the CST-based models have proved more successful than EOS-based models.

Generalized vapor pressure models typically require information about pure-fluid physical properties (such as the critical constants, normal boiling point, acentric factor, etc.). Often generalizations based on these traditional physical properties

* Corresponding author. Tel.: +1 405 744 5280; fax: +1 405 744 6338.
E-mail address: gasem@okstate.edu (K.A.M. Gasem).

0378-3812/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.fluid.2006.05.020

do not capture the subtleties of various chemical structures and thus produce poor predictions. Moreover, the necessary input properties are not available [11,12] for numerous new and/or challenging molecules (e.g., large hydrocarbons, hazardous chemicals). In some applications, group contribution (GC) methods provide reasonable predictions; however, here again, the first order approximation underlying GC methods compromises their abilities for many important fluids. As such, quantitative structure–property relationship (QSPR) models offer an attractive alternative since they have the potential to provide reliable property estimates based on *detailed* chemical structure information. Various literature studies which use structural descriptors for correlation of vapor pressure have been published; they include: (a) topological descriptors (e.g., molecular connectivity indices), (b) electronic descriptors (e.g., dipole moments and hydrogen bonding parameters), and (c) molecular parameters (e.g., molar volume and molar refractivity). Although successful in predicting vapor pressures for diverse organic molecules, currently available QSPR models [13–17] for vapor pressure are limited to predictions at a single temperature (298 K). The only comprehensive vapor pressure QSPR effort published to date which attempts to model data over the entire saturation range provides only semi-quantitative predictions, with an average error of 10% AAD [18]. Therefore, a vapor pressure model that provides accurate predictions (less than 1% AAD) over the entire saturation range, based solely on chemical structure information, is still needed.

In our previous work [19,20], we have developed a unified framework for correlating saturation properties, including vapor pressure. This scaled variable reduced coordinates (SVRC) model, which combines corresponding-states theory and scaling-law behavior, is capable of: (a) representing vapor pressures within their experimental uncertainties (generally within 0.1% AAD) using two fluid-specific regressed parameters, and (b) predicting vapor pressures with AAD of 1% using generalized parameters.

In this work, we generalize the SVRC model to a widely diverse set of organic molecules using structure–property modeling. Our primary goal is to develop a generalized model that predicts accurately vapor pressures of diverse organic molecules over the entire saturation range, -based solely on structural information. An extensive database of vapor pressure data involving 1121 molecules was used to develop this model. The approach calls for the use the QSPR methodology to generalize the model parameters of the previously developed SVRC model, thus, we use the SVRC to develop the behavior model, and QSPR to generalize the two substance-specific parameters of the model. This approach, we find, is more effective than the typical efforts to develop generalized models *directly* using QSPR techniques. The combined SVRC–QSPR model is more accurate and satisfies established theoretical limiting behavior of vapor pressure.

2. Vapor pressure prediction models

A review of vapor pressure models in the literature indicates that most find their origin in the Clausius–Clapeyron equation:

$$\frac{\mathrm{d}(\ln p)}{\mathrm{d}(1/T)} = -\frac{\Delta H_{\mathrm{lv}}}{R(\Delta Z_{\mathrm{lv}})} \tag{1}$$

where p is the pressure, T the temperature, and R is the gas constant. ΔH_{lv} and ΔZ_{lv} are changes in enthalpy and compressibility factor associated with vaporization. Integration of the Clausius–Clapeyron equation leads to most of the widely used vapor pressure correlations. However, this integration requires that assumptions be made regarding the temperature dependence of the $\Delta H_{\mathrm{lv}}/\Delta Z_{\mathrm{lv}}$ ratio. Assuming that $\Delta H_{\mathrm{lv}}/\Delta Z_{\mathrm{lv}}$ is constant leads to the simplest vapor pressure equation:

$$\ln p = A + \frac{B}{T} \tag{2}$$

where A is an integration constant and $B = -\Delta H_{\mathrm{lv}}/(R(\Delta Z_{\mathrm{lv}}))$. Both A and B are determined from experimental vapor pressure data. This correlation is precise in representing vapor pressures over small temperature intervals; however, it is imprecise when used to represent vapor pressure data over the entire saturation range.

Several modifications to this equation were proposed, these include: (a) the introduction of a third parameter into the denominator of Eq. (2) by Antoine [21]:

$$\ln p = A + \frac{B}{T + C} \tag{3}$$

and (b) the addition of two parameters C and D by Cox [22] to correct for deviation from the ideal gas law:

$$\ln p = A + \frac{B}{T} + CT + DT^2 \tag{4}$$

In Eq. (3), the constants A, B and C are substance specific parameters, which have been documented for over 700 pure fluids by Yaws and Yang [23]. Inspection of the parameter values indicates that, in general, the parameters A and B increase with carbon number, while C decreases. However, the use of this model has been limited due to its narrow applicable temperature and pressure range (1–200 kPa). Also, to date, no reliable generalized procedure for predicting these Antoine constants is available; as such, some experimental vapor pressure data are required for the purpose.

Maxwell and Bonnell [24] proposed a reference model approach to predict vapor pressures. They assumed that all molecules have a similar vapor pressure curve and used hexane as a reference. They were successful in predicting the vapor pressure curves of *n*-paraffins by using the hexane vapor pressure curve and two coefficients that are characteristics of each fluid. For other hydrocarbons they proposed a correction using the Watson characterization factor.

Correlations using two physical properties have been proposed by Gomez-Nieto and Thodos [25] and Iglesias-Silva et al. [26] While Gomez and Thodos used the boiling point and critical temperature for parameter generalization, Iglesias-Silva et al. developed an improved vapor pressure correlation using triple and critical point temperatures.

To increase the precision of vapor pressure representations, polynomial-type models with varying temperature exponents have also been proposed [6]:

$$\ln p_r = \frac{A(1-T_r) + B(1-T_r)^{1.5} + C(1-T_r)^3 + D(1-T_r)^6}{T_r} \quad (5)$$

where T_r and p_r are the reduced temperature and pressure, respectively. The model parameters *A–D* are empirical and experimental data are required to determine their values. Willman and Teja [27] proposed a set of equations for predicting the four coefficients of a modified Wagner equation as functions of the effective carbon number. However, these equations provide poor predictions if applied to fluids dissimilar to the molecules used for the correlation development. Other modifications to these equations have been proposed and the most widely cited works include Riedel [28], Thek and Stiel [29], Lee and Kesler [10], and Ambrose and Patel [30].

Correlations based on kinetic theory and group contributions have also been proposed. Abrams et al. [31] developed the following vapor pressure correlation based on the kinetic theory of fluids:

$$\ln p = A + \frac{B}{T} + C\ln T + DT + ET^2 \quad (6)$$

where the parameters *A–E* are expressed as functions of the number of equivalent oscillations per molecule, the enthalpy of vaporization of the hypothetical liquid at 0 K, and the hard-core van der Waals volume. For obtaining the hard-core van der Waals volume, Ambrose and Patel [30] recommended the group contribution method by Bondi [32]. The parameters *A* and *B* can be determined from known vapor pressure data, or alternatively they can be predicted by the group contribution method suggested by Macknick and Prausnitz [33]. Additional group contribution models have been proposed by Edwards and Prasunitz [34], Ruzicka [35], and Burkhard [36]. However most of these group contribution models are not suitable for predictions near the critical point. Also, correlations based on group contributions are mostly limited to first order approximations of the chemical structure and fail to provide accurate vapor pressure estimates when dealing with structural isomers.

Another approach that has gained significance over the years has been the use of chemical structure information to predict vapor pressure. Numerous QSPR efforts have been published in literature and include the works by Katritzky et al. [15], Goll and Jurs [3], Basaket al. [13], Beck et al. [14], Yaffe and Cohen [18] and Chalk et al. [37]. However, most of these efforts are limited to correlation of vapor pressure at a single temperature, 298 K. To date, no work has been published which correlates the vapor pressure over the entire saturation range (from the triple to critical point temperature) using QSPR.

3. Scaled variable reduced coordinate framework

In the 1980s, we developed the SVRC framework for correlating saturation properties [19,20]. The main characteristics of the framework are: (a) ability to correlate the saturation properties with high precision over the full saturation range, (b) ability to satisfy established scaling law behavior in the near-critical region, (c) suitability for generalization, which provides for a reliable predictive capability, and (d) simplicity. In later studies, the ability of the SVRC model to predict vapor pressures [11] and liquid [38] and vapor densities [39] of various fluids was investigated.

The general SVRC framework for correlation of thermophysical properties is given as [11,19,20]:

$$\frac{Y_\infty^\alpha - Y^\alpha}{Y_\infty^\alpha - Y_0^\alpha} = \Theta(\varepsilon) \quad (7)$$

or

$$Y^\alpha = Y_0^\alpha \Theta(\varepsilon) + [1-\Theta(\varepsilon)]Y_\infty^\alpha \quad (8)$$

where

$$\varepsilon = \frac{X_\infty - X}{X_\infty - X_0} \quad (9)$$

and $\Theta(\varepsilon)$ is the correlating function, X the correlating variable, Y the saturation property at given X, Y_∞ the asymptotic value of saturation property at X_∞, Y_0 the initial value of saturation property at X_0 and α is the scaling exponent.

For correlating saturation properties, Eq. (7) is recast for representing the various properties between the triple and the critical points as

$$Y = \{Y_c^\alpha - (Y_c^\alpha - Y_t^\alpha)\Theta\}^{(1/\alpha)} \quad (10)$$

Applied specifically to vapor pressure correlation, the above equation is written as

$$p = \{p_c^\alpha - (p_c^\alpha - p_t^\alpha)\Theta\}^{(1/\alpha)} \quad (11)$$

where p_c and p_t are the critical and triple point pressures, respectively. The functions $\Theta(\varepsilon)$ and α are defined as

$$\Theta(\varepsilon) = \frac{1 - A^{\varepsilon^B}}{1-A} \quad (12)$$

$$\alpha = \alpha_c - \Delta\alpha\frac{\varepsilon(1+C\varepsilon)}{1+C} \quad (13)$$

here

$$\varepsilon = \frac{T_c - T}{T_c - T_t} \quad (14)$$

and

$$\Delta\alpha = \alpha_c - \alpha_t \quad (15)$$

where *A*, *B* and *C* are universal correlation constants that apply to all fluids. Their values are 2/3, 0.985, and 4/3, respectively. The parameters α_c and α_t are the values of α at the critical and triple point temperatures, respectively.

In our SVRC framework, Eq. (7) accounts for the effects of temperature and chemical structure through the two functions (Θ and α), both of which are temperature dependent. Figs. 1 and 2 illustrate the variation of the reduced vapor pressure and scaled-and-reduced vapor pressure with reduced temperature, respectively. Inspection of the figures reveals that our SVRC model

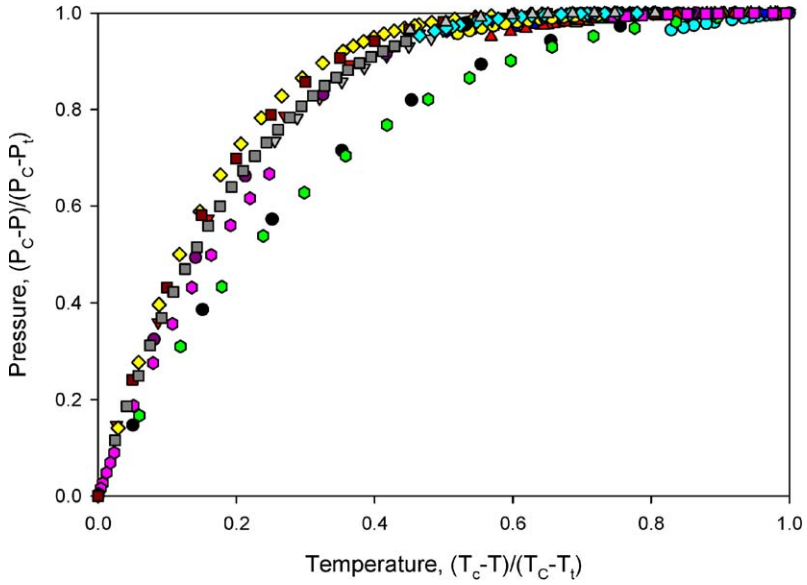

Fig. 1. Variation of reduced vapor pressure with temperature. (●) Alkanes; (▼) acetates; (■) acids; (◊) alcohols; (▲) aldehydes; (⬟) alkenes; (○) alkynes; (▽) amines; (■) anhydrates; (◆) cyanates; (▲) epoxides; (⬟) ethers; (⬟) formates; (▼) halogens; (■) ketones; (◆) mercaptans; (▲) multiring aromatics; (⬟) nitriles; (○) NO_2 compounds; (▼) peroxides; (■) polyols; (◇) sulphides; (△) terpenes.

effectively accounts for variations due to the chemical structure through the limiting values of the scaling exponent (α_c and α_t).

Generalizations of the two SVRC parameters, α_c and α_t, in terms of critical properties and acentric factor provided accurate vapor pressure predictions with an AAD of 1% for normal fluids (e.g., alkanes) and polar fluids (e.g., ammonia, refrigerants [38,39]). However, when the same generalization approach was extended to highly polar fluids such as alcohols, the model predictions were not as accurate [38] (2–5% AAD). We, therefore, sought an alternative to the traditional physical property-based parameter generalizations and explored the use of structure–property relationships. The preliminary QSPR generalization results obtained for 64 molecules indicated that this new SVRC–QSPR model provides better prediction results (1% AAD) compared to most literature models [14–18]. Therefore, in this work we extend the model to a diverse subset of over 1221 organic molecules containing most functional groups of importance to the chemical and environmental fields.

4. QSPR methodology

Several inherent problems challenge current QSPR model development strategies. Perhaps the most significant of these is the reliance on linear algorithms for identification of the

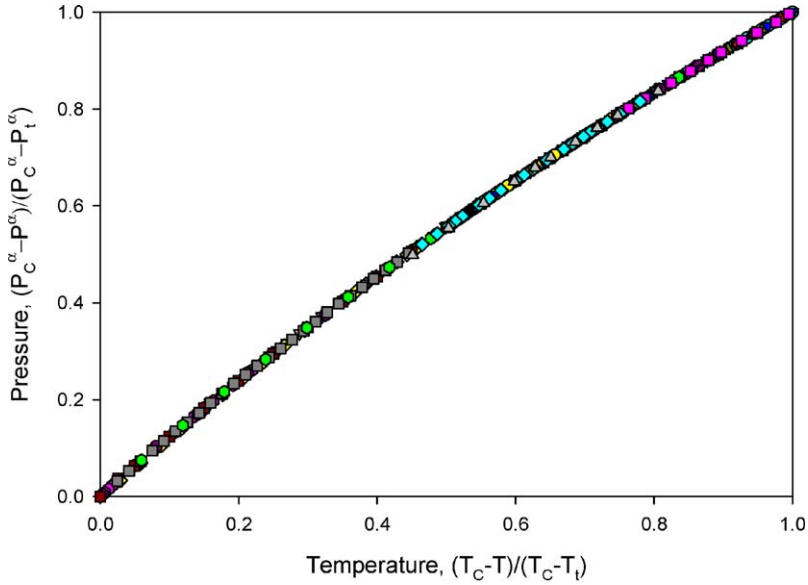

Fig. 2. Variation of scaled-and-reduced vapor pressure with temperature. (●) Alkanes; (▼) acetates; (■) acids; (◊) alcohols; (▲) aldehydes; (⬟) alkenes; (○) alkynes; (▽) amines; (■) anhydrates; (◆) cyanates; (▲) epoxides; (⬟) ethers; (⬟) formates; (▼) halogens; (■) ketones; (◆) mercaptans; (▲) multiring aromatics; (⬟) nitriles; (○) NO_2 compounds; (▼) peroxides; (■) polyols; (◇) sulphides; (△) terpenes.

significant structural descriptors. As stated earlier, the relationship between vapor pressure and chemical structure is non-linear; accordingly, non-linear algorithms have to be employed before reliable structure-based models can be evolved. However, commonly-used neural network architectures, such as back propagation networks [37], demand extensive training (using a significant amount of data) to develop a stable QSPR model. This places further demands on vapor pressure modeling efforts since reliable low-pressure experimental data are not plentiful.

In addition, vapor pressure has a non-linear temperature profile, as shown in Fig. 1. This type of functional behavior has not been addressed satisfactorily by current QSPR frameworks. Literature efforts which attempt to correlate directly vapor pressure through the use of structure–property relationship models or neural network models have only shown moderate success [18]. We, therefore, employ a theory-based non-linear model that provides reliable vapor pressure predictions using a moderate sampling of experimental data, and then generalize the model parameters using structure–property modeling. Specifically, we use the SVRC model provided in Eq. (11) to represent the vapor pressure data of each molecule using two substance-specific regressed parameters (α_c and α_t). These regressions provide a database of α_c and α_t values suitable for generalization. QSPR models are used to generalize these SVRC parameters in terms of the molecules' chemical structures. These generalized SVRC parameters are then used to predict the vapor pressures of the molecules.

QSPR models have been widely used in the literature to correlate the structure of a chemical with its thermo-physical behavior [40–43]. Following are the main elements of our vapor pressure QSPR model development effort: (a) assembly of a reliable vapor pressure database, (b) representation of the vapor pressure data using SVRC model, (c) generation of 3D structure of each molecule, (d) optimization of the 3D structures, (e) calculation of molecular descriptors, (f) reduction of the number of descriptors, (g) developing the generalized QSPR model for the SVRC parameters α_c and α_t, and (h) validation of the QSPR model. A flowchart outlining these processes is presented in Fig. 3.

4.1. Database development

The quality of the QSPR model development efforts depends on the availability of accurate experimental data. Our dataset of 1221 hydrocarbons was compiled from the Design Institute for Physical Property Data Project (DIPPR) 801 database [44]. However, since the database included vapor pressures from different data sources, a rigorous screening criterion was developed to determine the best subset of the data. Specifically, preference was given to data where: (a) the compound was a hydrocarbon, (b) the data were experimentally obtained, (c) had a DIPPR

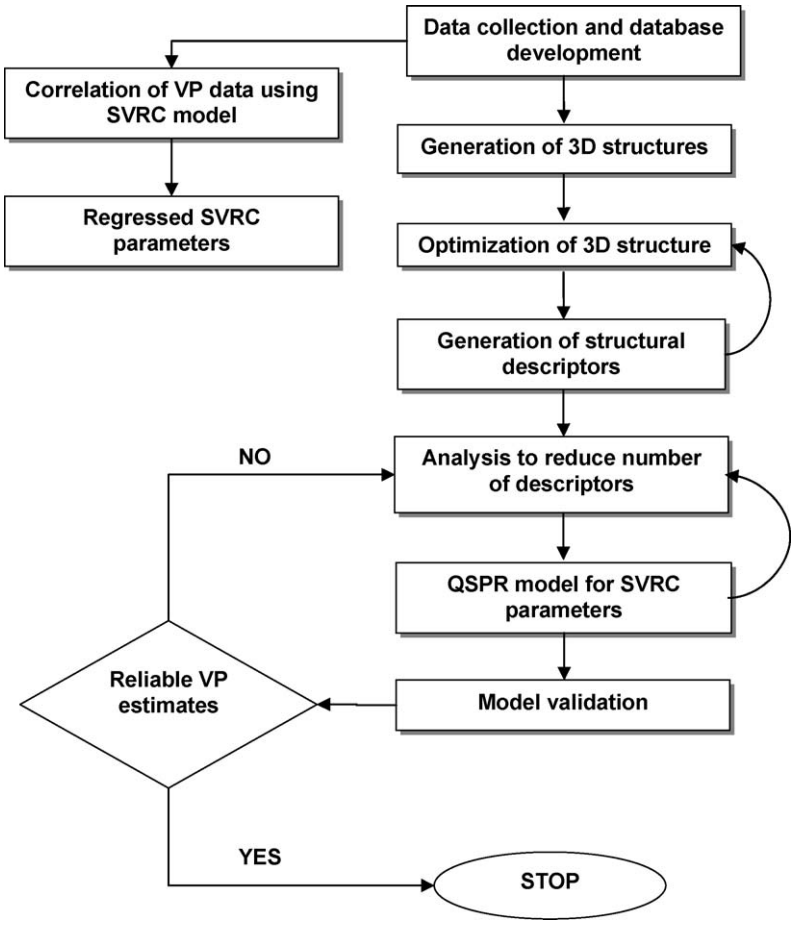

Fig. 3. Steps involved in development of a QSPR model.

quality rating of accepted, (d) had data over the entire saturation range, and (e) had at least six data points. A final data set containing 52,445 vapor pressure data points for 2691 data sets was used for model development. Table 1 presents a brief summary of the various classes of hydrocarbons, the number of molecules included and the number of datasets for each class of hydrocarbon. The complete list of molecules used in the development of the new model, along with the range of temperatures and pressure data obtained from DIPPR are provided in the supplementary material.

4.2. Generation and optimization of 3D structures

The 2D structure of each molecule was drawn using ChemDraw [45]. Several techniques for chemical representation, including simplified molecular input line system (SMILES) [46], 2D representation, connectivity tables [47] and 3D molecular coordinates [48], are in use today. Although all of the above-mentioned molecular representations find use to varying degrees, in this work, we use the 3D molecular representation. ChemDraw [45] software was used to convert the 2D structures into 3D structures. Further, to avoid inaccurate 3D molecule conformation, three different structure optimization routines (MOPAC [49], AMPAC [50] and GAUSSIAN [51]) were employed.

In general, more than one set of 3D coordinates (which satisfy the structural constraints of bond length and angle) can be generated for a given molecule, and it becomes necessary to locate the conformation with the minimum Gibbs free energy. Most chemicals contain single bonds that join two groups of atoms. Such bonds can usually rotate with a low-to-moderate energy barrier that changes the bond angle and orientation of other groups in the structure. Commercial structure optimization software typically uses Davidson–Fletcher–Powell (DFP) or Broyden–Fletcher–Goldfrab–Shannon (BFGS) algorithms as the default structure minimization algorithms. However, these routines fail if the initial input geometry is near a maximum [52]. Both DFP and BFGS try to minimize the potential energy surface and use a Hessian matrix to determine the search direction. The use of three optimization routines practically eliminated

Table 1
List of chemical classes included in this study and the number of molecules in each class

Class	Number of molecules	Number of Datasets	Class	Number of molecules	Number of datasets
1-Alkenes	20	79	*N*-alcohols	20	83
2,3,4-Alkenes	22	41	*N*-aliphatic acids	17	49
Acetates	16	45	*N*-aliphatic primary amines	13	35
Aldehydes	23	29	*N*-alkanes	31	121
Aliphatic ethers	24	50	*N*-alkylbenzenes	17	63
Alkylcyclohexanes	15	33	Naphthalenes	11	30
Alkylcyclopentanes	11	26	Nitriles	25	40
Alkynes	17	36	Nitroamines	4	4
Anhydrides	8	12	Organic salts	9	10
Aromatic alcohols	26	86	Other aliphatic acids	14	21
Aromatic amines	32	70	Other aliphatic alcohols	25	75
Aromatic carboxylic acids	9	10	Other aliphatic amines	19	29
Aromatic chlorides	14	21	Other alkanes	24	59
Aromatic esters	12	17	Other alkylbenzenes	37	113
C, H, BR compounds	16	26	Other amines, imines	27	32
C, H, F compounds	31	134	Other condensed rings	9	18
C, H, I compounds	8	12	Other ethers/diethers	19	28
C, H, multihalogen compounds	35	109	Other hydrocarbon rings	11	19
C, H, NO_2 compounds	19	27	Other monoaromatics	16	26
C1/C2 aliphatic chlorides	18	51	Other polyfunctional C, H, O	24	31
C3 and higher aliphatic chlorides	21	31	Other polyfunctional organics	4	4
Cycloaliphatic alcohols	7	8	Other saturated aliphatic esters	10	20
Cycloalkanes	6	32	Peroxides	7	8
Cycloalkenes	10	15	Polyfunctional acids	14	14
Dialkenes	23	45	Polyfunctional amides/amines	21	25
Dicarboxylic acids	13	13	Polyfunctional C, H, N, halide (O)	8	10
Dimethylalkanes	21	49	Polyfunctional C, H, O, halide	28	32
Diphenyl/polyaromatics	15	25	Polyfunctional C, H, O, N	23	26
Epoxides	11	28	Polyfunctional C, H, O, S	8	9
Ethyl and higher alkenes	12	27	Polyfunctional esters	10	13
Formates	12	25	Polyfunctional nitriles	2	2
Isocyanates/diisocyanates	7	8	Polyols	27	46
Ketones	31	100	Propionates and butyrates	12	24
Mercaptans	21	50	Sulfides/thiophenes	26	62
Methylalkanes	17	53	Terpenes	7	10
Methylalkenes	22	44	Unsaturated aliphatic esters	14	21
Multiring cycloalkanes	3	12			

Table 2
Examples of molecular descriptors calculated from QSPR packages

Constitutional
Number of atoms
Molecular weight
Geometrical
Moments of inertia
Molecular volume
Thermodynamic
Vibrational enthalpy
Internal entropy at 300 K
Topological
Wiener index
Kier shape index
Electrostatic
Charged partial surface area
Min and max partial charges
Quantum-chemical
LUMO
HOMO

un-optimized structures. In general, the commercial structure optimization packages available today provide an accurate estimate of the 3D coordinates for *most* molecules. In the few cases where all the routines fail to converge; we typically re-optimize the structure after introducing a small change in the structure of the molecule.

4.3. Descriptor generation

Traditionally, QSPR model development has relied on the use of single commercial software for descriptor generation (CODESSA [53], ADAPT [54], DRAGON [55]). Our experience has been that descriptor generation programs require continuous updating. As such, the use of descriptors from single software, although attractive due to the reduced computational requirement, does not assure model stability. In this study, we employed all three commercial packages which provide a total of 1400 molecular descriptors per molecule, including a variety of constitutional, topological, and electrostatic descriptors. The actual number of descriptors calculated for each molecule varied based on the structural complexity of the molecule. Examples from each group of descriptors are presented in Table 2.

4.4. Descriptor reduction

Not all the 1400 descriptors generated using commercial QSPR software are significant to modeling of vapor pressures. The use of all available descriptors in model development causes dimensionality problems. Further, use of irrelevant or redundant descriptors diminishes the performance of a QSPR model, especially when non-linear algorithms are used in the model development. Initially, orthogonalization algorithms are used to obtain a subset of descriptors from which all redundant information has been removed. Two types of orthogonalization schemes were employed in this work.

The first approach [56] involves the calculation of inter-correlations for all descriptors involved. The descriptors generated for molecules are usually represented as vectors. If descriptor $|k\rangle$ is being tested for orthogonality against descriptor $|j\rangle$ and r_{jk} is the correlation coefficient, then descriptor $|k\rangle$ is retained if the r_{jk} value is low, otherwise it is eliminated. This process is repeated for all descriptors. In the second approach [57], the most significant descriptors were identified using linear regression analysis and then the descriptors are orthogonalized. If descriptor $|j\rangle$ is the first descriptor whose correlation with the property is measured, and if the second descriptor $|k\rangle$ is obtained from the linear analysis but exhibits high inter-correlation with descriptor $|j\rangle$, then the second descriptor is mathematically transformed by using Eq. (16):

$$\left|k'\right\rangle = \frac{r_{jk}\,|j\rangle - |k\rangle}{[1 - r_{kj}^2]^{1/2}} \tag{16}$$

where $\left|k'\right\rangle$ is the orthogonalized $|k\rangle$ descriptor, while r_{jk} is the correlation coefficient between the non-orthogonal descriptors $|j\rangle$ and $|k\rangle$. The same procedure is repeated for all the remaining descriptors.

The orthogonalization techniques alone are incapable of identifying the optimal set of descriptors, and hundreds of descriptors may remain after orthogonalization; thus, further reduction and fine-tuning of these descriptors before model development is required. For this study, linear and non-linear descriptor reduction algorithms were used to identify the most significant subset of descriptors. Several methods for descriptor reduction have been validated previously in our laboratories for QSPR model development, including principal component analysis (PCA) [58], partial least squares (PLS) [59], genetic algorithms (GAs) [60], and neural networks (NNs) [61]. However, in this work descriptor reduction using only neural networks was found adequate. A detailed description of the methodology can be found elsewhere [38,39,56,62,63].

5. QSPR model development

The primary objectives of this work were: (a) to demonstrate the ability of the SVRC model to accurately represent vapor pressure using two regressed parameters, (b) to examine the efficacy of using existing QSPR algorithms to obtain estimates for SVRC parameters based on chemical structure information alone, and (c) using non-linear neural network algorithms to improve SVRC parameter predictions. To attain these objectives, three case studies were conducted sequentially:

Case 1. Determination of SVRC parameters through regression of experimental data.
Case 2. Generalization of SVRC parameters using linear structure-based models.
Case 3. Generalization of SVRC parameters using non-linear structure-based models.

The vapor pressure model development was initiated with the regression of SVRC parameters for each molecule (case 1).

As can be seen from Eqs. (12)–(15), the SVRC model for vapor pressure has five parameters A, B, C, α_c, α_t. However, previous studies [19,20] have indicated that excellent model precision is obtained when the parameters A, B, and C are treated as universal constants applicable to all fluids. Further, when α_c and α_t are treated as substance specific, the suitability of the two-parameter SVRC model to generalization (based on traditional physical properties, e.g., critical properties and acentric factor) has already been demonstrated [11]. However, these generalizations fail when extended to molecules with challenging functional groups. Clearly, a more rigorous generalization strategy based on chemical-structure is required.

Structure-based model development is typically initiated with linear analysis. In this work: (a) multi-linear regression analysis and (b) heuristic analysis were used to identify the most significant subset of molecular descriptors. However, vapor pressure exhibits a non-linear relationship between structural descriptors and model parameters. Therefore, the use of linear relationships to model such behavior often yields poor predictions. Although valuable insight is gained from the linear modeling, inevitably, robust non-linear algorithms are required before meaningful model generalizations can be obtained.

In this study, the best subset of descriptors identified through linear analysis was used to develop the non-linear QSPR models (case 2). Then, a simple back-propagation (BP) network was employed for non-linear model development (case 3). Several BP algorithms have been proposed and are in use for prediction of thermo-physical properties. Two widely used BP algorithms are: (a) the gradient descent and (b) the Levenberg–Marquardt algorithms [64]. In this work, the Levenberg–Marquardt algorithm was employed. The development of a BP network involves the use of a trial-and-error approach for selection of architecture and network weights. Networks differ in the number of layers and hidden neurons used in each layer. Usually each network

Table 3
Summary of vapor pressure results obtained using regressed and predicted SVRC parameters

Class	%AAD					Class	%AAD				
	Reg	Gen	Linear	NL	NL*		Reg	Gen	Linear	NL	NL*
1-Alkenes	0.12	5.96	1.01	0.27	0.13	*N*-alcohols	0.31	9.56	1.04	0.55	0.24
2,3,4-Alkenes	0.4	5.56	1.22	0.47	0.29	*N*-aliphatic acids	0.38	12.19	1.74	0.57	0.23
Acetates	0.25	7.12	0.99	0.41	0.27	*N*-aliphatic primary amines	0.43	7.82	0.94	0.59	0.36
Aldehydes	0.49	7.62	1.52	0.53	0.34	*N*-alkanes	0.26	5.15	1.24	0.40	0.17
Aliphatic ethers	0.34	4.57	1.31	0.45	0.29	*N*-alkylbenzenes	0.26	6.96	0.98	0.35	0.15
Alkylcyclohexanes	0.22	4.98	0.84	0.29	0.18	Naphthalenes	0.30	4.91	1.46	0.41	0.25
Alkylcyclopentanes	0.14	8.23	0.68	0.19	0.08	Nitriles	0.41	15.18	1.37	0.47	0.30
Alkynes	0.39	5.24	1.22	0.49	0.31	Nitroamines	0.58	5.59	1.35	0.58	0.37
Anhydrides	0.61	8.48	1.34	0.66	0.33	Organic salts	0.55	35.19	0.96	0.55	0.29
Aromatic alcohols	0.3	7.47	1.20	0.49	0.22	Other aliphatic acids	0.69	11.00	2.75	0.80	0.53
Aromatic amines	0.29	6.08	1.29	0.41	0.21	Other aliphatic alcohols	0.27	9.23	1.52	0.44	0.27
Aromatic carboxylic acids	0.17	8.53	0.50	0.19	0.11	Other aliphatic amines	0.40	5.47	1.08	0.45	0.26
Aromatic chlorides	0.32	7.99	0.79	0.38	0.24	Other alkanes	0.16	5.57	0.78	0.19	0.08
Aromatic esters	0.38	25.66	1.38	0.45	0.19	Other alkylbenzenes	0.20	5.77	1.34	0.28	0.17
C, H, BR compounds	0.33	13.27	0.93	0.37	0.20	Other amines, imines	0.41	4.65	2.70	0.43	0.26
C, H, F compounds	0.16	1.92	1.01	0.3	0.15	Other condensed rings	0.26	4.99	0.85	0.35	0.16
C, H, I compounds	0.53	9.39	0.96	0.59	0.24	Other ethers/diethers	0.60	5.72	1.77	0.65	0.34
C, H, multihalogen compounds	0.18	2.47	0.99	0.32	0.15	Other hydrocarbon rings	0.12	2.99	0.45	0.14	0.09
C, H, NO_2 compounds	0.62	12.17	1.59	0.68	0.41	Other monoaromatics	0.35	10.08	1.06	0.42	0.28
C1/C2 aliphatic chlorides	0.26	4.01	1.25	0.44	0.21	Other polyfunctional C, H, O	0.50	19.35	1.95	0.55	0.35
C3 and higher aliphatic chlorides	0.22	6.14	0.98	0.29	0.17	Other polyfunctional organics	0.62	44.20	1.34	0.62	0.41
Cycloaliphatic alcohols	0.4	8.38	1.05	0.46	0.32	Other saturated aliphatic esters	0.27	12.29	1.31	0.42	0.22
Cycloalkanes	0.05	1.00	0.79	0.18	0.12	Peroxides	0.42	12.63	1.10	0.45	0.20
Cycloalkenes	0.63	4.31	1.49	0.68	0.29	Polyfunctional acids	0.36	14.83	0.70	0.36	0.18
Dialkenes	0.44	5.00	1.49	0.52	0.33	Polyfunctional amides/amines	0.38	7.17	1.56	0.41	0.23
Dicarboxylic acids	0.45	18.21	0.96	0.45	0.18	Polyfunctional C, H, N, halide (O)	0.45	4.29	1.58	0.48	0.29
Dimethylalkanes	0.2	4.23	0.58	0.25	0.16	Polyfunctional C, H, O, halide	0.34	5.32	1.38	0.39	0.19
Diphenyl/polyaromatics	0.31	3.59	0.88	0.38	0.22	Polyfunctional C, H, O, N	0.27	6.04	1.10	0.30	0.13
Epoxides	0.23	6.76	1.01	0.35	0.16	Polyfunctional C, H, O, S	0.61	9.97	1.90	0.62	0.26
Ethyl and higher alkenes	0.34	4.39	0.77	0.39	0.22	Polyfunctional esters	0.68	6.46	2.44	0.73	0.45
Formates	0.64	5.69	1.88	0.77	0.45	Polyfunctional nitriles	0.17	2.81	0.20	0.17	0.11
Isocyanates/diisocyanates	0.52	4.19	1.42	0.53	0.26	Polyols	0.47	38.05	2.07	0.57	0.23
Ketones	0.26	6.40	1.17	0.4	0.27	Propionates and butyrates	0.24	5.76	0.88	0.36	0.18
Mercaptans	0.16	5.81	0.75	0.23	0.16	Sulfides/thiophenes	0.11	8.55	1.00	0.20	0.13
Methylalkanes	0.28	8.25	1.66	0.38	0.25	Terpenes	0.72	7.70	2.14	0.75	0.32
Methylalkenes	0.35	6.34	1.05	0.43	0.22	Unsaturated aliphatic esters	0.52	6.30	1.75	0.58	0.31
Multiring cycloalkanes	0.36	4.11	0.99	0.59	0.26	Overall	0.36	7.44	1.24	0.44	0.24

NL*: predictions obtained when the normal point was used as the lower vapor pressure point to predict the vapor pressure between the normal boiling point and the critical point.

architecture and weight combination is tried out at least once before deciding the best layout.

In some cases, retraining of the network might be needed if the model predictions are poor. The poor predictions occur mostly if the network is used to extrapolate, i.e., when the network is used to predict properties of molecules which contain functional groups that are not present in the training set. Therefore, for reliable predictions, one must make sure that the training set encompasses all the possible functional groups that might be encountered in the design phase. Also of importance is the handling of outliers in the training data because outliers can affect significantly the performance of the network. Over fitting [65] is the other significant problem with BP networks. When trained on the same data set for multiple cycles, the network develops a surface that accommodates all the training data; this provides an accurate estimate of the training set but has poor predictive capabilities. Increasing the number of hidden neurons also decreases the ability of a given network to predict. Typically for a BP network the number of hidden neurons is between one and the number of descriptors being used in the model. There are several heuristics that provide guidance as to the permissible number of hidden neurons or choice of networks weights and criteria for termination of training. For example, when determining the network architecture, the rule of thumb is to keep the ratio of total observations (number of molecules) to adjustable parameters (network weights) above two to avoid any chance effects [66]. Nevertheless, these rules are approximations and trial-and-error investigation remains the reliable method for choosing the best parameters.

After rigorous trial-and-error analysis, a 10-12-1 network architecture was obtained; specifically, 10 descriptors were identified for each of the SVRC parameters and used as network inputs, and 12 hidden neurons were used for network prediction. Two separate 10-12-1 networks, one each for the SVRC parameters (α_c and α_t), were developed. To validate the neural network model, the data set containing 1221 molecules was split randomly into training, prediction and cross validation (CV) subsets. The prediction set and CV set each contained 20% of the original data while the training set contained the other 60%. The network was trained using only the training data. Periodically, the training was halted and property values were predicted for the CV set. The root mean square error (RMSE) of this prediction was plotted as a function of the number of training cycles. The RMSE of the training set decreases as the number of training cycles increase [42,43], while the RMSE in the CV set passes through a minimum. The minimum in the CV set indicates where the training should be halted. Validation of the predictive ability of the non-linear neural network model was conducted using the prediction set. The SVRC parameters for the vapor pressure are predicted using the final neural network obtained after training and cross validation. If the RMSE error for the prediction set is comparable to those of the training and CV sets, then the network model is accepted [42,43]. The network weights are chosen at random, thus multiple initializations of the network are required before a stable model can be evolved. For this vapor pressure model development, over 100 random network initializations were used to identify the best network architecture.

Table 4
Descriptors obtained from linear modeling of SVRC vapor pressure parameter α_c

Descriptors	Coefficients
FPSA-3 (Zefirov's PC)	−6.74E−01
Reduced temperature (T_r)	3.29E−01
Min net atomic charge for a H atom	1.13E−01
Acentric factor	−9.35E−02
Triple point pressure (p_t)	1.70E−02
Kier flexibility index	2.74E−03
Translational entropy	−4.90E−04
Moment of inertia C	3.09E−04
DPSA-2 (Zefirov's PC)	−2.35E−04
Boiling point (T_B)	−1.06E−04
Intercept	2.35E−01

6. Results and discussion

The entire data set containing over 1121 molecules was correlated using the two parameter SVRC model. The DIPPR database included vapor pressure data from more than one source for some molecules; in such cases, each data source was treated as an individual dataset resulting in a total of 2961 datasets.

Table 3 presents the summary results for the case studies conducted, and the supplementary materials provide a listing of the SVRC parameters obtained through regression and prediction. As indicated by the results for case 1, the SVRC model represents the vapor pressures of the diverse organic subset over the full saturation range, on average, within 0.36% using two regressed parameters. These results confirm our previous findings [38,39,67] on the efficacy of the SVRC in mathematically mapping the vapor pressure behavior of various chemical species.

The two SVRC parameters were then generalized using linear QSPR models (case 2). Multi-linear regression analysis was used and the parameter regressions that provided the best parameter predictions were retained. Tables 4 and 5 present the selected descriptors identified through linear analysis. While the linear model provided satisfactory estimates for the SVRC parameters, large errors (%AAD >2) were noticed for vapor pressure predictions. A sensitivity analysis indicated that an error of 1% in prediction of SVRC parameters translated into an error of 2.6%

Table 5
Descriptors obtained from linear modeling of SVRC vapor pressure parameter α_t

Descriptors	Coefficients
Reduced temperature (T_r)	5.38E−01
Max partial charge for a Cl atom	−2.90E−01
Max nucleoph. react. index for a C atom	2.58E−01
Max net atomic charge for a H atom	−1.77E−01
Acentric factor	−1.33E−01
XY shadow/XY rectangle	3.42E−02
Triple point pressure (p_t)	4.69E−03
Randic index (order 1)	−4.11E−03
Moment of inertia C	6.70E−04
HOMO–LUMO energy gap	4.11E−04
Intercept	2.55E−03

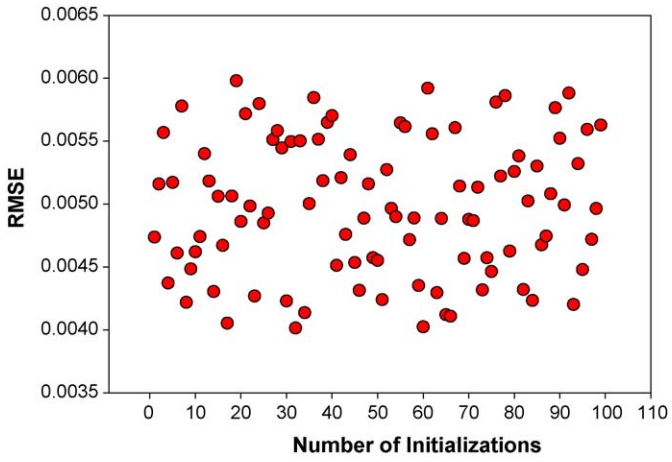

Fig. 4. Plot of RMSE obtained for 100 random network weight initializations.

AAD in vapor pressure. However, the best linear model was able to predict the SVRC parameters only with 1.2% AAD. While this prediction accuracy is acceptable for most applications, we sought a more robust non-linear algorithm that could provide more accurate predictions.

The descriptors identified through linear analysis were used to develop the non-linear neural network model (case 3). A three-layer, feed forward network trained with the back-propagation of errors and a sigmoidal transfer function were used in this study. Several architectures for the hidden neurons from 6 to 16 hidden neurons were investigated. The 10-12-1 network architecture was found to provide the best performance as measured from the cross validation set error.

The neural networks were initiated, as typically done, with a set of network weights chosen at random. The network progressively altered these weights as training proceeds. However, since the network performance is sometimes dependent on the initial weights used, multiple network training sessions using different initial weights are required to assure model reliability. Thus, 100 separate networks with random starting weights were trained with randomly selected training and prediction sets. Training of the neural networks was halted when the RMSE of the cross-validated predictions for the entire dataset reached a minimum. Fig. 4 provides the network RMS error values obtained for each of the hundred network initializations.

We set our target for a robust model to have a maximum error of twice the average experimental uncertainty. Analysis of our preliminary prediction results for case 3 indicated the existence of several data points with large errors (>5%). All datasets (396 datasets) which had at least a single data point with large prediction error were isolated and subjected to further analysis. The source of the error in these datasets was determined to be the choice of critical properties used. All available saturated vapor pressure data were included in our model development and as such, our data includes experimental data from multiple authors for most molecules. When the same critical property data were used for all datasets, poor predictions were obtained for a few systems. Previously, Span and Wagner [68] had discussed the influence of errors in critical temperature on model stability and had suggested that critical temperatures could vary by as much as 3% depending on the apparatus and operating procedures. When the 396 systems were regressed allowing for a 3% variation in critical temperature, a significant reduction in the error was noticed. Therefore, all critical property data provided by DIPPR were examined further using external data sources *a priori* to the subsequent model development.

The prediction results for a single subset of molecules (alkanes) are shown graphically in Figs. 5 and 6. Fig. 5 shows the deviation in vapor pressure predicted using generalized SVRC parameters while Fig. 6 provides a comparison plot of experimental and predicted vapor pressures for alkanes. On average,

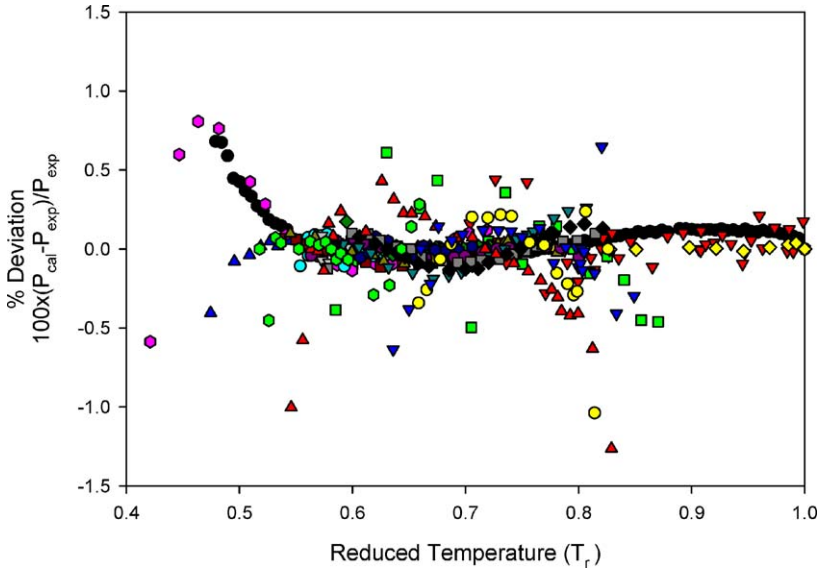

Fig. 5. Deviations in predicted vapor pressures for alkanes obtained using QSPR generalized SVRC parameters. (●) Methane; (▼) ethane; (■) propane; (◊) butane; (▲) pentane; (⬟) hexane; (⬟) heptane; (▽) octane; (■) nonane; (◆) decane; (▲) undecane; (⬟) dodecane; (⬟) tridecane; (▼) tetradecane; (■) pentadecane; (◆) hexadecane; (▲) heptadecane; (⬟) octadecane; (⬟) nonadecane; (▼) eicosane.

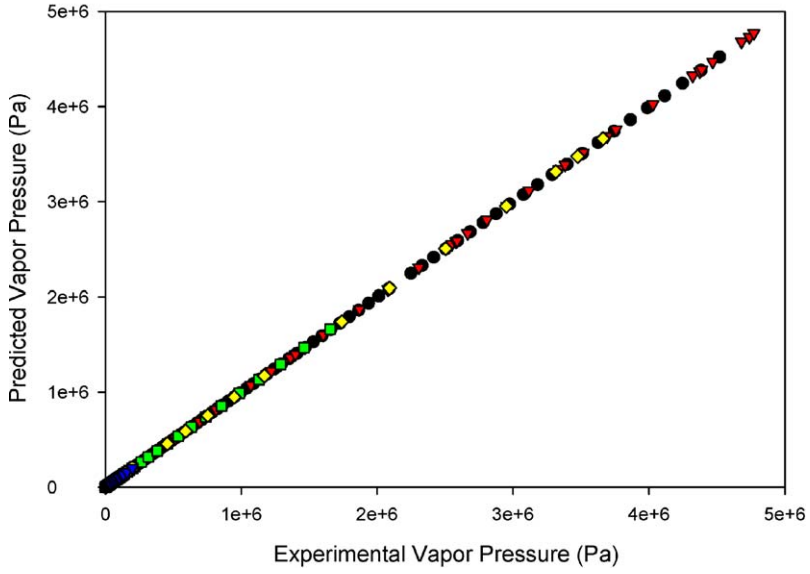

Fig. 6. Comparison of experimental and predicted vapor pressures for alkanes obtained using QSPR generalized SVRC parameters. (●) Methane; (▼) ethane; (■) propane; (◊) butane; (▲) pentane; (⬟) hexane; (⬮) heptane; (▽) octane; (■) nonane; (◆) decane; (▲) undecane; (⬟) dodecane; (⬟) tridecane; (▽) tetradecane; (■) pentadecane; (◆) hexadecane; (▲) heptadecane; (⬟) octadecane; (◯) nonadecane; (▼) eicosane.

the SVRC model was capable of representing the vapor pressure data for the diverse organic subset within 0.4% AAD and predicting within 0.5% AAD. To obtain a truly generalized model the entire dataset was divided into a training, prediction and cross validation datasets. Only a small fraction of the entire data was used for training. The smallest and largest molecule from each of the 73 classes of hydrocarbons was included in the training set. Overall, of the 2691 datasets available, only 645 datasets were used for training, while the rest were used for prediction and cross validation. Molecules used for cross validation were randomly chosen from each of the classes of hydrocarbons accounting for a total of 283 datasets.

Of the 1763 datasets used for prediction only five datasets exhibited large prediction errors (greater than 2% AAD) and include: (a) 9-octadecanoic acid (2.75% AAD), (b) hexahydroazepine (2.70% AAD), (c) 2-hydroxy benzoic acid methyl ester (2.44% AAD), (d) 1-methyl-4-(1-methylethylidine) cyclohexane (2.14% AAD), and (e) 2,2′-(1,2-ethanediylbis(oxy))bisethanol (2.07% AAD). Of the remaining datasets, 94 systems had error greater than 1%, and 500 systems had error greater than 0.5%. Aliphatic acids (for example methylhexanoic acid) produced the highest average error of 0.8% AAD, while nitriles and hydrocarbons with rings yielded the least prediction errors. The overall distribution of prediction errors for the entire data set can be seen in Fig. 7.

When the triple point data are not available as input parameters, we have found that the normal point (or the lowest available temperature) may be used as the lower vapor pressure point to predict the vapor pressure between the normal boiling point and the critical point with good accuracy. Prediction results using the normal boiling point as the lower end point are given in Table 3. As indicated by the overall AAD of 0.24%, the quality of the predictions for most systems has improved. These results may be attributed to the fact that vapor pressure measurements below the atmospheric pressure contain larger deviations.

Our goal in this study was to develop a generalized vapor pressure model capable of providing accurate *a priori* predictions: (a) over the full saturation range from the triple point to the critical point, and (b) for all systems, including those for which no experimental data are available. Our SVRC model provides superior vapor pressure representations (0.3% AAD for 1221 molecules) in comparison to any of the recommended literature models such as the two-reference-fluid model or the Gomez–Thodos model [69]. Also the our SPR–SVRC model outperforms any published structure-based generalization study, including the works published by Katritzky et al. [15], Wessel and Jurs [42,43], and Yaffe and Cohen [18]. In fact, we believe that prior to this work the use of QSPR models to represent temperature dependent properties had not been adequately demonstrated; only moderate success was achieved (AAD of 10%) by using the traditional QSPR approaches [18].

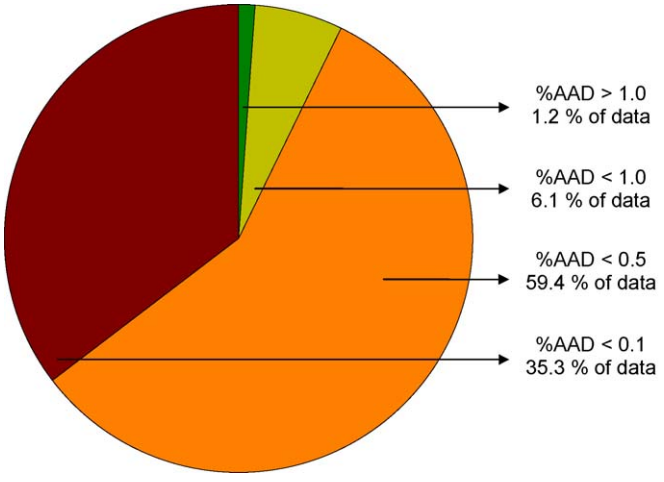

Fig. 7. Distribution of errors in predicted vapor pressures obtained using the generalized SVRC–QSPR model.

The basic assertion of this study is that neither EOS models nor neural network-based QSPR models alone provide satisfactory vapor pressure predictions for diverse chemical structures. Thus, we hypothesized that a two-pronged approach, which calls for the use of theory to develop the behavior model, and QSPR to generalize the parameters of such model, is more effective. The quality of the vapor pressure predictions obtained for such a diverse group of molecules (73 chemical classes involving 52,445 data points) demonstrates the validity of this integrated approach and provides credible evidence in support of our hypothesis.

7. Conclusions

1. Our SVRC framework successfully correlated vapor pressures of a diverse subset of organic molecules from the triple to the critical point. The SVRC model was able to represent vapor pressures of 1221 molecules within 0.4% AAD on average when two adjustable parameters were used for each substance.
2. A 10-12-1 back-propagation network was developed to generalize the SVRC parameters (α_c, α_t) based on structural descriptors. The SVRC–QSPR models provide excellent generalized predictions of vapor pressures, with average errors of less than 0.5%, based on triple point and critical point data and structural descriptors.
3. The model predictions improved further (0.24% AAD) when predictions were restricted to temperatures above the boiling point.
4. The results of this study indicate that the use of theory-framed structure–property modeling is effective in thermo-physical property model generalization.

References

[1] T.E. Daubert, D.K. Jones, Project 821: pure component liquid vapor pressure measurements, AIChE Symp. Ser. 86 (1990) 29–39.
[2] M. Reinhard, A. Drefahl, Handbook for Estimating Physicochemical Properties of Organic Compounds, Wiley & Sons, New York, 1999.
[3] E.S. Goll, P.C. Jurs, Prediction of vapor pressures of hydrocarbons and halohydrocarbons from molecular structure with a computational neural network model, J. Chem. Inf. Comput. Sci. 39 (1999) 1081–1089.
[4] H.W. Xiang, Vapor pressures from a corresponding-states principle for a wide range of polar molecular substances, Int. J. Thermophys. 22 (2001).
[5] D.J. Chase, Two models for vapor pressure along the saturation curve, Ind. Eng. Chem. Res. 26 (1987) 107.
[6] W. Wagner, New vapor pressure measurements for argon and nitrogen, and a new method for establishing rational vapor pressure equation, Cryogenics 13 (1973) 470.
[7] W. Waring, Form of a wide range vapor pressure equation, Ind. Eng. Chem. Res. (1954) 762.
[8] T. Charoensombut-amon, R.J. Marti, R. Kobayashi, Applications of generalized multiproperty apparatus to measure phase equilibrium and vapor phase densities of supercritical carbon dioxide in *n*-hexadecane systems up to 26 MPa, Fluid Phase Equilib. 31 (1986) 89.
[9] K.A.M. Gasem, C.H. Ross, R.L. Robinson Jr., Prediction of ethane and CO_2 solubilities in heavy paraffins using generalized-parameter Soave and Peng-Robinson equations of state, Can. J. Chem. Eng. 71 (1993) 805.
[10] B.I. Lee, M.G. Kesler, A generalized thermodynamic correlation based on three-parameter corresponding states, AIChE J. 21 (1975) 510.
[11] R.D. Shaver, R.L. Robinson Jr., K.A.M. Gasem, A framework for the prediction of saturation properties: vapor pressures, Fluid Phase Equilib. 64 (1991) 141–163.
[12] G.R. Somayajulu, New vapor pressure equations from triple point to critical point and predictive procedure for vapor pressure, J. Chem. Eng. Data 31 (1986) 438.
[13] S.C. Basak, B.D. Gute, G.D. Grunwald, Use of topostructural, topochemical, and geometric parameters in the prediction of vapor pressure: a hierarchical QSAR approach, J. Chem. Inf. Comput. Sci. 37 (1997) 651–655.
[14] B. Beck, A. Breindl, T. Clark, QM/NN QSPR models with error estimations: vapor pressure and Log *P*, J. Chem. Inf. Comput. Sci. 40 (2000) 1046–1051.
[15] A.R. Katritzky, Y. Wang, S. Sild, T. Tamm, M. Karelson, QSPR studies on vapor pressure, aqueous solubility, and the prediction of water–air partition coefficients, J. Chem. Inf. Comput. Sci. 38 (1998) 720–725.
[16] C. Liang, D.A. Gallagher, QSPR prediction of vapor pressure from solely theoretically-derived descriptors, J. Chem. Inf. Comput. Sci. 38 (1998) 321–324.
[17] H.E. McClelland, P.C. Jurs, Quantitative structure–property relationships for the prediction of vapor pressures of organic compounds from molecular structures, J. Chem. Inf. Comput. Sci. 40 (2000) 967–975.
[18] D. Yaffe, Y. Cohen, Neural network based temperature-dependent quantitative structure–property relations (QSPRs) for predicting vapor pressure of hydrocarbons, J. Chem. Inf. Comput. Sci. 41 (2001) 463–477.
[19] K.A.M. Gasem, J.R.L. Robinson, R.D. Shaver, Phase equilibrium data in development of correlations in coal fluids: a new framework for correlating saturation properties, #1 vapor pressure predictions, Department of Energy Progress Report, DEFG22-83PC60039-11, 1988.
[20] K.A.M. Gasem, J.R.L. Robinson, R.D. Shaver, Phase equilibrium data in development of correlations in coal fluids: a new framework for correlating saturation properties, #2 liquid density predictions, Department of Energy Progress Report, DEFG22-83PC60039-11, 1988.
[21] C. Antoine, Tensions des vapeurs: nouvelle relation entre les tensions et les températures, C. R. Acad. Sci. 107 (1888) 681.
[22] E.R. Cox, Pressure–temperature chart for hydrocarbon vapors, Ind. Eng. Chem. 15 (1923) 592.
[23] C.L. Yaws, H.-C. Yang, To estimate vapor pressure easily, Hydrocarbon Proc. 68 (1989) 65.
[24] J.B. Maxwell, L.S. Bonnell, Derivation and precision of a new vapor pressure correlation for petroleum hydrocarbons, Ind. Eng. Chem. 49 (1957) 1187.
[25] M. Gomez-Nieto, G. Thodos, A new vapor pressure equation and its application to normal alcanes, Ind. Eng. Chem. Fundam. 16 (1977) 2.
[26] G.A. Iglesias-Silva, J.C. Holste, P.T. Eubank, K.N. Marsh, K.R. Hall, A vapor pressure equation from extended asymptotic behavior, AIChE J. 33 (1987) 1550.
[27] B. Willman, A.S. Teja, Method for the prediction of pure component vapor pressures in the range 1 kPa to the critical pressure, Ind. Eng. Chem. Process Des. Dev. 24 (1985) 1033–1036.
[28] L. Riedel, A New Universal Formula for Vapour Pressure, Chem. Eng. Technol. 26 (1954) 83.
[29] R.E. Thek, L.I. Stiel, A new reduced vapor pressure equation, AIChE J. 12 (1966) 599–602.
[30] D. Ambrose, N.C. Patel, J. Chem. Thermodyn. 16 (1984) 459–468.
[31] D.S. Abrams, H.A. Massaldi, J.M. Prausnitz, Vapor pressures of liquids as a function of temperature. Two-parameter equation based on kinetic theory of fluids, Ind. Eng. Chem. Fundam. 13 (1974) 259.
[32] A.A. Bondi, Physical Properties of Molecular Crystals, Liquids, and Glasses, John Wiley & Sons, New York, 1968.
[33] A.B. Macknick, J.M. Prausnitz, Vapor pressures of heavy liquid hydrocarbons by a group-contribution method, Ind. Eng. Chem. Fundam. 18 (1979) 348.
[34] D.R. Edwards, J.M. Prausnitz, Estimation of vapor pressures of heavy liquid hydrocarbons containing nitrogen and sulfur by a group-contribution method, Ind. Eng. Chem. Fundam. 20 (1981) 280.
[35] V. Ruzicka Jr., Estimation of vapor pressures by a group-contribution methodology, Ind. Eng. Chem. Fundam. 22 (1983) 266.

[36] L.P. Burkhard, Estimation of vapor pressures for halogenated aromatic hydrocarbons by a group-contribution method, Ind. Eng. Chem. Fundam. 24 (1985) 119.
[37] A.J. Chalk, B. Beck, T. Clark, A temperature-dependent quantum mechanical/neural net model for vapor pressure, J. Chem. Inf. Comput. Sci. 41 (2001) 1053–1059.
[38] S.S. Godavarthy, D. Ravindranath, R.L. Robinson Jr., K.A.M. Gasem, Generalized SVRC–QSPR predictions of saturated vapor pressure and phase densities, in: AIChE Spring National Meeting, New Orleans, LA, 2004.
[39] S.S. Godavarthy, D. Ravindranath, K.A.M. Gasem, R.L. Robinson Jr., Generalized SVRC–QSPR predictions of saturated vapor densities, in: AIChE Spring Meeting, Atlanta, GA, 2005.
[40] L.M. Egolf, M.D. Wessel, P.C. Jurs, Prediction of boiling points and critical temperatures of industrially important organic compounds from molecular structure, J. Chem. Inf. Comput. Sci. 34 (1994) 947.
[41] A.R. Katritzky, U. Maran, M. Karelson, V.S. Lobanov, Prediction of melting points for the substituted benzenes: a QSPR approach, J. Chem. Inf. Comput. Sci. 37 (1997) 913.
[42] M.D. Wessel, P.C. Jurs, Prediction of normal boiling points of hydrocarbons from molecular structure, J. Chem. Inf. Comput. Sci. 35 (1995) 68.
[43] M.D.J. Wessel, P.C. Jurs, Prediction of normal boiling points for a diverse set of industrially important organic compounds from molecular structure, J. Chem. Inf. Comput. Sci. 35 (1995) 841.
[44] Design Institute for Physical Property Data, Project 801, Physical and Thermodynamic Property Database, 1999.
[45] Chemdraw 8. 0, Cambridge Software, 2004.
[46] D. Weininger, SMILES introduction and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1988) 31.
[47] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[48] M. Randic, M. Razinger, The characterization of three-dimensional molecular structure, in: A.T. Balaban (Ed.), Chemical Topology to Three-Dimensional Geometry, Plenum Press, New York, 1996, pp. 159–236.
[49] J.P.P. Stewart, Mopac: a semi-empirical orbital program, J. Comput. Aided Mol. Des. 4 (1990) 1–103.
[50] AMPAC 6. 0, Semichem Inc., Shawnee, KS, 1998.
[51] M.J. Dewar, E.G. Zoebisch, E.F. Healy, J.J.P. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, J. Am. Chem. Soc. 107 (1985) 898–3902.
[52] X. Zhou, R. Liu, P. Pulay, Theoretical study of the structure, force field and vibrational spectra of cyclooctatetraene, Spectrochim. Acta A 49 (1993) 953.
[53] CODESSA 2. 63, Semichem Inc., Shawnee, KS, 1998.
[54] ADAPT, P.C. Jurs, University Park, PA, 1998.
[55] DRAGON 4. 0, Milano Chemometrics, Italy, 2003.
[56] M. Karelson, Molecular Descriptors in QSAR/QSPR, John Wiley & Sons, New York, 2000.
[57] J.M. Sutter, T.A. Peterson, P.C. Jurs, Prediction of gas chromatographic relative retention times of alkylbenzenes, Anal. Chim. Acta 342 (1997) 113–122.
[58] E.R. Malinowski, D.G. Howery, Factor Analysis in Chemistry, John Wiley & Sons, New York, 1980.
[59] S. Wold, P. Geladi, K. Esbensen, J. Ohman, J. Chemometrics 1 (1987).
[60] D. Rogers, A.J. Hopfinger, Application of generic function approximation to quantitative structure–activity relationships and quantitative structure–property relationships, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[61] J. Zupan, J. Gasteiger, Neural Networks for Chemists, VCH Publishers, Weinheim, 1993.
[62] S.S. Godavarthy, Design of improved solvents for extractive distillation, PhD Thesis, Oklahoma State University, 2004.
[63] D. Ravindranath, Structure-based generalized models for pure-fluid saturation properties and activity coefficients, MS Thesis, Oklahoma State University, 2005.
[64] T.M. Mitchell, Machine Learning, McGraw-Hill, New York, 1997.
[65] A.A. Gakh, E.G. Gakh, B.G. Stumper, D.W. Noid, Neural network-graph theory approach to the prediction of the physical properties of organic compound, J. Chem. Inf. Comput. Sci. 34 (1994) 832–839.
[66] C.W. Ulmer, D.A. Smith, B.G. Sumpter, D.I. Noid, Computational neural networks and the rational design of polymeric materials: the next generation polycarbonates, Comput. Theor. Polym. 8 (1998) 311–321.
[67] C.J. Schult, D.W. Towmey, R.L. Robinson Jr., K.A.M. Gasem, Prediction of vapor pressures of refrigerants using the SVRC correlating framework, in: Third International Petroleum Environmental Conference, Albuquerque, NM, 1996.
[68] R. Span, W. Wagner, A new equation of state for carbon dioxide covering the fluid region from the triple point temperature to 1100 K and pressures up to 800 MPa, J. Phys. Chem. Ref. Data 25 (1996) 1509–1590.
[69] R.C. Reid, J.M. Prausnitz, J. O'Connell, The Properties of Gases and Liquids, 5th ed., McGraw-Hill, New York, 2000.