

Development of an Improved Group Contribution Method for the Prediction of Vapour Pressures of Organic Compounds

By

Bruce Moller [B.Sc.(Eng.)]

University of KwaZulu-Natal Durban

In fulfilment of the degree Master of Science (Chemical Engineering)

December 2007

ABSTRACT

Vapour pressure is an important property in the chemical and engineering industries. There are therefore many models available for the modelling of vapour pressure and some of the popular approaches are reviewed in this work. Most of the more accurate methods require critical property data and most if not all require vapour pressure data in order to regress the model parameters. It is for this reason that the objective of this work was to develop a model whose parameters can be predicted from the molecular structure (*via* group contribution) or are simple to acquire via measurement or estimation (which in this case is the normal boiling point).

The model developed is an extension of the original method that was developed by Nannoolal et al. The method is based on the extensive Dortmund Data Bank (DDB), which contains over 180 000 vapour pressure points (for both solid and liquid vapour pressure as of 2007). The group parameters were calculated using a training set of 113 888 data points for 2332 compounds. Structural groups were defined to be as general as possible and fragmentation of the molecular structures was performed by an automatic procedure to eliminate any arbitrary assumptions. As with the method of Nannoolal the model only requires knowledge about the molecular structure and the normal boiling point in order to generate a vapour pressure curve. In the absence of experimental data it is possible to predict the normal boiling point, for example, by a method developed by Nannoolal et al.

The relative mean deviation (RMD) in vapour pressure was found to be 5.0 % (2332 compounds and 113 888 data points) which compares very well with the method of Nannoolal et al. (6.6 % for 2207 compounds and 111 757 data points). To ensure the model was not simply fitted to the training set a test set of liquid vapour pressure, heat of vaporization and solid vapour pressure data was used to evaluate its performance. The percentage error for the test set was 7.1 % for 2979 data points (157 compounds). This error is artificially high as the test data contained a fair amount of less reliable data. For the heat of vaporization at 298.15 K (which is related to vapour pressure via the Clausius-Clapeyron equation) the RMD was 3.5 % for 718 compounds and in the case of solid vapour pressures the RMD error was 21.1 % for 4080 data points (152 compounds). Thus the method was shown to be applicable to data that was not contained in the training set.

PREFACE

The work presented in this dissertation was undertaken at the University of KwaZulu-Natal Durban from January 2007 till December 2007. The work was supervised by Professor D. Ramjugernath and Professor Dr. J. Rarey.

This dissertation is presented as the full requirement for the degree of Master of Science in Engineering (Chemical). All work presented in this dissertation is original unless otherwise stated and has neither in whole or part been submitted previously to any tertiary institute as part of a degree.

Bruce Moller (203502126)

As the candidate's supervisor I have approved this dissertation for submission

Prof. D. Ramjugernath

ACKNOWLEDGEMENTS

I wish to acknowledge the following people and organizations for their contribution to this work:

- My supervisors, Prof. D. Ramjugernath and Prof. Dr. J. Rarey for tireless support, ideas and motivation during this work. Working with them has been a privilege and a great learning experience.
- Dr. Y. Nannoolal for helping me to get this work started.
- NRF International Science Liaison, NRF-Thuthuka Programme and BMBF (WTZ-Project) for financial support.
- DDBST GmbH for providing data and software support for this project.
- My parents Erik and Jeanette, my sister Teresa, my grandmother and late grandparents for all their years of wisdom and guidance. They have my eternal gratitude.
- Most importantly I would like to give praise to my Lord and Saviour Jesus Christ. “... *Praise be to the name of God for ever and ever; wisdom and power are his. He changes times and seasons; he sets up kings and deposes them. He gives wisdom to the wise and knowledge to the discerning.*” Daniel 2:20-21.

TABLE OF CONTENTS

Abstract	i
Preface	ii
Acknowledgements	iii
Table of Contents	iv
List of Figures	viii
List of Tables	xiii
Nomenclature	xvii
1 Introduction	1
2 Theory and Literature Review	3
2.1 Introduction	3
2.2 Vapour pressure models.....	4
2.2.1 Classical thermodynamics.....	4
2.2.1.1 The Antoine equation.....	8
2.2.1.2 The Cox equation and Cox charts	10
2.2.1.3 The Riedel equation.....	12
2.2.1.4 The Myrdal & Yalkowsky equation.....	14
2.2.1.5 The Tu group contribution method.....	16

2.2.1.6	The modified Watson equation	19
2.2.2	Kinetic theory of vaporization	20
2.2.3	Equations of state	22
2.2.3.1	Alpha functions	22
2.2.3.2	Lee-Kesler method.....	25
2.2.4	Empirical models	27
2.2.4.1	The Wagner equation	28
2.2.4.2	Quantitative structure property relationship	29
2.2.4.3	Interpolation polynomials	30
2.3	Solvation theory	31
2.4	Solid vapour pressures	32
3	Computation and Database tools.....	33
3.1	Database.....	33
3.2	Data validation	33
3.3	Regression.....	35
3.3.1	Linear regression.....	35
3.3.2	Non-linear regression	36
3.3.3	Inside-Outside regression.....	39

3.3.4	Implementation	40
3.3.5	Fragmentation	41
4	Development of the method.....	43
4.1	Model development	43
4.2	The group contribution concept	48
4.3	The group interaction concept	48
4.4	New group contribution approach.....	49
5	Results and discussion	50
5.1	Hydrocarbon compounds.....	50
5.2	Oxygen compounds.....	56
5.3	Nitrogen compounds.....	63
5.4	Sulfur compounds	67
5.5	Halogen compounds.....	68
5.6	Other compounds	69
5.7	Testing the method	69
5.8	Solid vapour pressures	70
5.9	Heat of vaporization	71
5.10	Solubility parameters	72

5.11	Advantage of group contribution	73
5.12	General results and discussion	74
6	Conclusions.....	77
7	Recommendations	78
8	References.....	79
	Appendices	84
A	Group contribution and interaction tables.....	84
B	Sample calculations.....	94
C	Riedel calculation example.....	96
D	Equations for $\Delta H/R\Delta Z$	97
E	Software screenshots.....	98
F	Calculation of ΔH_{vap} from equations of state	100
G	Change in the chemical potential	104
H	Further notes on data validation and data used.....	105

LIST OF FIGURES

Figure 2.1 Phase diagram for water (semi log plot) – The solid liquid equilibrium (SLE) data points are only illustrative and not experimental data.....	3
Figure 2.2 Heat of vaporization of benzene as a function of temperature (♦ – data from the DDB, — Watson equation [Eq. (2-36) with $m = 0.391$])	7
Figure 2.3 ΔZ_{vap} of benzene as a function of temperature using the SRK EOS with Twu alpha function (Twu et al.).....	7
Figure 2.4 $\Delta H_{\text{vap}}/\Delta Z_{\text{vap}}$ of benzene as a function of reduced temperature ($T_r = T/T_c$) (calculated from the Watson and the SRK EOS using the Twu alpha function (Twu et al.)).....	8
Figure 2.5 $\ln(P^S/101.3 \text{ kPa})$ vs. $1/T$ for benzene showing Antoine plots fitted to different temperature ranges (x – data taken from the DDB, — 270 K to 560 K, - - - 270 K to 300 K, — 350 K to 380 K, — 380 K to 410 K)	9
Figure 2.6 Antoine prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - calculated from SRK and the Watson equation for benzene, — Antoine prediction – Eq. (D-1))	10
Figure 2.7 Cox prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - calculated from SRK and the Watson equation for benzene, — Cox prediction – Eq. (D-2))	12
Figure 2.8 Riedel prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Riedel prediction – Eq. (D-3), - - - Riedel direct fit – Eq. (D-3)).....	14
Figure 2.9 Best possible Mydral & Yalkowsky prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Mydral & Yalkowsky prediction – Eq. (D-4)).....	16
Figure 2.10 Best possible Tu prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Tu prediction – Eq. (D-5))	18
Figure 2.11 Best possible “modified Watson” prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — “Modified Watson” prediction – Eq. (D-6)).....	20

Figure 2.12 Abrams et al. prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Abrams et al. prediction – Eq. (D-7)).....	22
Figure 2.13 SRK prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — SRK prediction – Appendix F)	25
Figure 2.14 Lee-Kesler prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Lee-Kesler prediction – Eq. (D-8))	27
Figure 2.15 Wagner prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Wagner prediction – Eq. (D-9)).....	29
Figure 2.16 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for benzene, with solid vapour pressure data (x – liquid data taken from the DDB, x – solid data taken from the DDB — solid vapour pressure, - - - - sub cooled liquid vapour pressure).....	32
Figure 3.1 Experimental data from the DDB for amyl formate	34
Figure 3.2 Experimental data from the DDB for n-eicosane	34
Figure 3.3 Flow diagram for the “Inside-Outside” regression technique	40
Figure 3.4 The group definition and structure of the aliphatic carboxylic acid group.....	41
Figure 4.1 A proper fit for the Eq. (2-27)	43
Figure 4.2 A physically unrealistic fit for Eq. (2-27).....	43
Figure 4.3 $\ln(P^s/1\text{ kPa})$ vs. T for 1-butanol (♦ - data from the DDB, — with the logarithm term, - - - - without the logarithm term).....	45
Figure 4.4 $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ for 1-butanol (♦ - data from SRK using the MC alpha function and the Watson equation (Eq. (2-36) with $m = 0.473$), — prediction with the logarithm term, - - - - prediction without the logarithm term).....	45
Figure 4.5 B' vs. polarizability for the n-alkanes	46

Figure 4.6 B' vs. polarizability for hydrocarbons	47
Figure 4.7 T_b vs. polarizability for hydrocarbons	47
Figure 4.8 Comparison of the group contribution approaches for the non-cyclic alkene groups.....	49
Figure 5.1 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for propane (x – data taken from the DDB, - - - - data regressed using the C-parameter correlation of Nannoolal et al., ——— data regressed using the improved C-parameter correlation)	51
Figure 5.2 P^s vs. T for propane (x – data taken from the DDB, - - - - data regressed using the C-parameter correlation of Nannoolal et al., ——— data regressed using the improved C-parameter correlation)	52
Figure 5.3 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for octadecane (x – data taken from the DDB, - - - - data regressed using the C-parameter correlation of Nannoolal et al., ——— data regressed using the improved C-parameter correlation)	52
Figure 5.4 P^s vs. T for octadecane (x – data taken from the DDB, - - - - data regressed using the C-parameter correlation of Nannoolal et al., ——— data regressed using the improved C-parameter correlation)	53
Figure 5.5 Comparison of the properties of anthracene and phenanthrene.....	53
Figure 5.6 dB_i vs. number of atoms for different alkynes (♦ – dB_i data for each compound, — a linear least squares fit)	54
Figure 5.7 dB_i vs. number of atoms for different alkenes (♦ – dB_i data for each compound, — a linear least squares fit)	55
Figure 5.8 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for 1-nonanol (x – data taken from the DDB, - - - - data regressed using the method of Nannoolal et al., ——— data regressed using the new logarithmic correction)	57
Figure 5.9 P^s vs. T for 1-nonanol (x – data taken from the DDB, - - - - data regressed using the method of Nannoolal et al., ——— data regressed using the new logarithmic correction)	57

Figure 5.10 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for palmitic acid (x – data taken from the DDB, - - - data regressed using the method of Nannoolal et al., — data regressed using the new logarithmic correction)	58
Figure 5.11 P^s vs. T for palmitic (x – data taken from the DDB, - - - data regressed using the method of Nannoolal et al., — data regressed using the new logarithmic correction)	58
Figure 5.12 dB_i vs. number of atoms for different epoxides (♦ – dB_i data for each compound, — a linear least squares fit)	59
Figure 5.13 dB_i vs. number of atoms for different ketones (♦ – dB_i data for each compound, — a linear least squares fit)	60
Figure 5.14 dB_i vs. number of atoms for different aliphatic alcohols (♦ – dB_i data for each compound, — lines to show the trends)	60
Figure 5.15 dB_i vs. number of atoms for different aliphatic carboxylic acids (♦ – dB_i data for each compound, — lines to show the trends)	61
Figure 5.16 GI_j vs. number of atoms for different diols (♦ – dB_i data for each compound, — a linear least squares fit)	61
Figure 5.17 dB_i vs. number of atoms for different primary aliphatic amines (♦ – dB_i data for each compound, — a linear least squares fit)	64
Figure 5.18 dB_i vs. number of atoms for different nitriles (♦ – dB_i data for each compound, — lines to show the trends)	64
Figure 5.19 dB_i vs. number of atoms for different aliphatic isocyanates (♦ – dB_i data for each compound, — a linear least squares fit)	65
Figure 5.20 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for thymol (x – solid data taken from the DDB, — vapour pressure curve)	71
Figure 5.21 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for diethyl malonate (x – data taken from the DDB, — predicted, - - - fitted)	74

Figure 5.22 Histogram of the vapour pressure relative mean deviation for the compounds in the training set.....	75
---	----

Appendices

Figure E.1 Screenshot of the GUI used to validate the data – this enabled the fast removal of any obvious outliers	98
--	----

Figure E.2 Screenshot of the GUI used to test the different models – enabled the rapid testing of the forms of the equation that were tested for this work	99
--	----

Figure F.1 P vs. V for water at 560 K as given by the van der Waals EOS.....	101
--	-----

Figure F.2 Heat of vaporization of benzene as a function of temperature (♦ – data from the DDB, — SRK prediction Eq. (F-13)).....	103
---	-----

Figure F.3 Heat of vaporization of benzene as a function of temperature (♦ – data from the DDB, — Twu SRK prediction Eq.(F-14) and Eq. (F-9) in the form of Eq. (F-13))	103
---	-----

LIST OF TABLES

Table 2.1 Antoine constants for various temperature ranges	9
Table 2.2 Relative mean deviations (RMD) for vapour pressures of selected compounds – Antoine equation.....	10
Table 2.3 Relative mean deviations (RMD) for vapour pressure of selected compounds – Cox equation.....	11
Table 2.4 Relative mean deviations (RMD) for vapour pressure of selected compounds – Riedel equation.....	13
Table 2.5 Relative mean deviations (RMD) for vapour pressure of selected compounds - Mydral & Yalkowsky equation.....	15
Table 2.6 Relative mean deviations (RMD) for vapour pressure of selected compounds – Tu equation.....	18
Table 2.7 Relative mean deviations (RMD) for vapour pressure of selected compounds – Eq. (2-38)	20
Table 2.8 Relative mean deviations (RMD) for vapour pressure of selected compounds – method of Abrams et al	22
Table 2.9 Relative mean deviations (RMD) for vapour pressure of selected compounds – SRK EOS using Eqn. (2-48).....	25
Table 2.10 Relative mean deviations (RMD) for vapour pressure of selected compounds – Lee-Kesler	26
Table 2.11 Relative mean deviations (RMD) for vapour pressure of selected compounds – Wagner (* - parameters fitted by author)	29
Table 5.1 New hydrocarbon structural groups (Ink No – fragmentation group number, Ref No –	

reference number is used to arrange like groups (e.g. halogen groups etc) since the ink no's have no real structure) 50

Table 5.2 Relative mean deviation [%] in vapour pressure estimation for different types of hydrocarbons (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 55

Table 5.3 Relative mean deviation [%] in vapour pressure estimation for different types of hydrocarbons (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 56

Table 5.4 Relative mean deviation [%] in vapour pressure estimation for different types of oxygen containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 62

Table 5.5 Relative mean deviation [%] in vapour pressure estimation for different types of oxygen containing compounds (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 63

Table 5.6 Relative mean deviation [%] in vapour pressure estimation for different types of nitrogen containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 66

Table 5.7 Relative mean deviation [%] in vapour pressure estimation for different types of nitrogen containing compounds (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of

compounds; ELP – Extremely low pressure $P < 10 \text{ Pa}$; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 66

Table 5.8 Relative mean deviation [%] in vapour pressure estimation for different types of sulfur containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10 \text{ Pa}$; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 67

Table 5.9 Relative mean deviation [%] in vapour pressure estimation for different types of sulfur containing compounds (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10 \text{ Pa}$; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 67

Table 5.10 Relative mean deviation [%] in vapour pressure estimation for different types of halogen containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10 \text{ Pa}$; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 68

Table 5.11 Relative mean deviation [%] in vapour pressure estimation for different types of halogen containing compounds (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10 \text{ Pa}$; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 68

Table 5.12 Relative mean deviation [%] in vapour pressure estimation for various other compounds. The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10 \text{ Pa}$; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error. 69

Table 5.13 Relative mean deviation [%] in vapour pressure estimation for various other compounds (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.....	69
Table 5.14 Percentage errors for the test set data	70
Table 5.15 Percentage errors for heat of vaporization at 298.15 K using Eq. (5-2) with $\Delta Z = 1$	72
Table 5.16 Results for the prediction of Hildebrand solubility parameters at 298 K	73
Table 5.17 Relative mean deviation [%] in vapour pressure estimation for the new method and the method of Nannoolal et al. The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.	74

Appendices

Table A.1 Group contribution and group interaction values and descriptions	84
Table A.2 Group contribution values for the logarithmic correction term.....	93
Table B.1 Calculation for the vapour pressure of 1-hexen-3-ol at 389K.....	94
Table B.2 Calculation for the vapour pressure of 2-mercapto ethanol at 364.8 K.....	95
Table D.1 Forms of the equations for $\Delta H/R\Delta Z$ for the various equations used in this dissertation ..	97

NOMENCLATURE

a, b, c, d	-	model parameters/constants
A,B,C,D,E	-	model parameters/constants
B',D'	-	parameters predicted by group contribution
C _p	-	constant pressure heat capacity (J/mol.K)
G	-	Gibbs free energy (J/mol)
H	-	enthalpy (J/mol)
M	-	molar mass (g/mol)
n	-	number of moles
P	-	pressure (kPa)
R	-	ideal gas constant (J/mol.K)
S	-	entropy (J/mol.K)
T	-	absolute temperature (K)
V	-	molar volume (cm ³ /mol)
Z	-	compressibility factor
Greek symbols		
μ	-	chemical potential
ω	-	Pitzer acentric factor

Subscripts

atm	-	atmospheric
b	-	normal boiling point
c	-	critical point
m	-	normal melting point
r	-	reduced property
vap	-	vaporization

Superscripts

s	-	saturated / solid
l	-	liquid
v	-	vapour

All other symbols used are explained in the text and unless otherwise stated SI units have been used.

1 INTRODUCTION

Vapour pressure has for a long time been an important property in chemical and engineering applications. It is useful in the design of distillation columns, storage and transport of materials and for determining cavitation in pumps to name a few examples. It is also important for predicting the fate of chemicals in the environment due to its predominant effect on the distribution coefficient between air and various other compartments (e.g. air and water). Daubert¹ ranked vapour pressure second, behind critical properties, in the list of the most important thermophysical properties (ranking is based on required accuracy and uses).

Many fitted (i.e. fitted directly to data) and predictive methods are available for the representation of the vapour pressure curve. Correlated (or fitted) methods are usually good over the range of data fitted but some extrapolate very poorly if not fitted to a wide enough data range (some of these methods will be discussed in the following chapters). A drawback of fitted models is that their parameters require experimental data which in many cases are not available. This means that a suitable quantity of the chemical (if not readily available) must be synthesized and vapour pressure measurements undertaken.

Even though the Dortmund Data Bank (DDB²) contains over 180 000 data points for more than 6000 compounds, it is only a fraction of the more than 100 000 (according to the environmental news network³) chemicals that are reported to be in use today. This coupled with the fact that new chemicals are continually being discovered (Bowen et al.⁴ estimate 200 to 1000 chemicals per annum) means that measurement is not only very expensive but also impractical (even though there is such a large amount of chemicals being discovered per annum many of these chemicals have vapour pressure which is too low to be of practical concern). For this reason accurate prediction methods have become increasingly popular.

A popular approach for the prediction of thermophysical properties is group contribution methods. The component is broken down into structural groups (e.g. CH₃, OH etc.). Their contributions are combined to describe the behaviour of the whole molecule. The methods are especially popular for properties like boiling point and critical properties, but surprisingly few exist (or are published) for vapour pressure prediction.

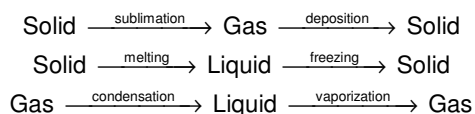
In the preceding work of Rarey et al.⁵ and Nannoolal et al.^{6,7,8}, group contribution estimation methods for the normal boiling point, critical data and vapour pressure of organic non-electrolyte compounds were presented. The objective of this work is to extend and improve the method for

vapour pressure estimation. This was achieved by addition of more data to the training set, further critical examination of the training set data and extended utilization of low pressure data for higher molecular weight components. Structural and functional groups were defined in such a way as to make the method as widely applicable as possible. Due to the importance of vapour pressure data the predictions should be reasonably accurate (usually within 5%) and have a low probability of total failure (i.e. errors in excess of 15%).

2 THEORY AND LITERATURE REVIEW

2.1 Introduction

By adding or removing heat from a pure substance or changing the system pressure, one can change the phase of the substance. Some of the common phase transitions are as follows:



These phase transitions are often represented on a diagram known as a phase diagram. Figure 2.1 shows a phase diagram for water, the solid lines are the phase boundaries or the lines of equilibrium between the phases. In addition to the phases shown there are often different phases in the solid region. If the type of solid phase changes along the sublimation curve, a discontinuity in the slope of the coexistence curve is observed.

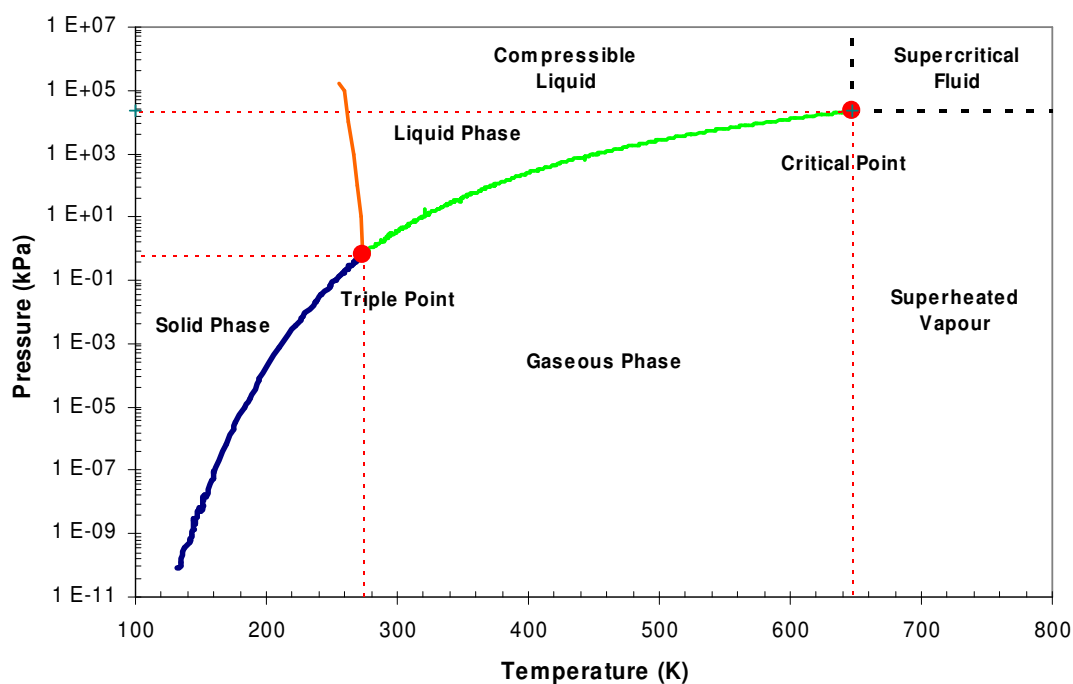


Figure 2.1 Phase diagram for water (semi log plot) – The solid liquid equilibrium (SLE) data points are only illustrative and not experimental data.

The lines of equilibrium show where 2 phases coexist, and the triple point is where solid/liquid/vapour all coexist. Consider a liquid in a sealed container with a vapour space above the liquid; the molecules in the vapour phase will eventually reach a state of dynamic equilibrium, with the rate of vaporization being equal to the rate of condensation. The vapour space is then said to be saturated and the resulting pressure in the container is called the saturated vapour pressure.

The boiling point of a substance is defined as the temperature at which the saturated vapour pressure is equal to the ambient pressure. The most common vapour pressure point is the normal boiling point, it is the temperature at which the saturated vapour pressure is 1 atm (this is known as the standard atmospheric pressure and is defined to be 101.325 kPa).

Due to its importance in process simulation (specifically distillation), vapour pressure is regarded as one of the most important thermophysical properties. Daubert¹ ranks vapour pressure as the second most important thermophysical property, whereby his ranking system is based on the use of the property on its own, its input into other equations and the accuracy to which the property should be known. Unsurprisingly critical properties were ranked number one mainly due to the large number of corresponding states methods and correlations that are based on this reference point (many of the more accurate vapour pressure correlations also use critical data).

Many equations have been developed to describe the vapour pressure from the triple point to the critical point. The Wagner⁹ equation has been shown to be able to reproduce the curve but it requires knowledge of the critical point and accurate data. Therefore an equation is required which gives the correct behaviour where only few data are available. Many of the vapour pressure equations that are used in industry today have their roots in the Clausius-Clapeyron equation.

2.2 Vapour pressure models

2.2.1 Classical thermodynamics

A thermodynamic treatment of the pure component phase equilibrium described above was presented by Gibbs and further refined by other researchers (esp. Riedel, Ambrose). Gibbs introduced a quantity known as the chemical potential. The chemical potential of a species i , is given by the change in the total Gibbs free energy of a system if one mole (or molecule) of this species is removed or added. This process must not alter the state of the system, therefore the chemical potential is defined as:

$$\mu_i = \left(\frac{\partial G}{\partial n_i} \right)_{T,P,n_{j \neq i}} \quad (2-1)$$

At a particular temperature and pressure the phase which has the lower chemical potential will be the more stable phase. Taking the example of the container used above, if the temperature of the liquid is suddenly raised the following will result (see Appendix G):

$$\mu^v < \mu^l \quad (2-2)$$

This means that there will be transfer of mass from the liquid to the vapour phase until the chemical potentials in both phases are equal. Therefore for all points on the equilibrium curve the following holds (α and β represent the 2 phases on the curve):

$$\mu^\alpha = \mu^\beta \quad (2-3)$$

Since for a pure substance the chemical potential is only a function of 2 of the 3 (there is no composition) state variables, we chose T and P (since we want an expression for vapour pressure) and take the total differential of both sides of Eq. (2-3) is:

$$\left(\frac{\partial \mu^\alpha}{\partial T} \right)_P dT + \left(\frac{\partial \mu^\alpha}{\partial P} \right)_T dP = \left(\frac{\partial \mu^\beta}{\partial T} \right)_P dT + \left(\frac{\partial \mu^\beta}{\partial P} \right)_T dP \quad (2-4)$$

The differential form of the Gibbs function is (G,S and V are all molar properties):

$$dG = -SdT + VdP \quad (2-5)$$

Eq. (2-5) has the same form as Eq. (2-4) and since $G = \mu$ (for a pure substance) the following two expressions arise:

$$\left(\frac{\partial \mu^i}{\partial P} \right)_T = V^i \quad (2-6)$$

$$\left(\frac{\partial \mu^i}{\partial T} \right)_P = -S^i \quad (2-7)$$

Substituting Eq. (2-6) and Eq. (2-7) into Eq. (2-4) yields:

$$-S^{\alpha}dT + V^{\alpha}dP = -S^{\beta}dT + V^{\beta}dP \quad (2-8)$$

This can be rewritten as:

$$\frac{dP}{dT} = \frac{S^{\alpha} - S^{\beta}}{V^{\alpha} - V^{\beta}} \quad (2-9)$$

Since the two phases are in equilibrium, Eq. (2-10) holds and substituting this into Eq. (2-9) yields Eq.(2-11):

$$S^{\alpha} - S^{\beta} = \frac{H^{\alpha} - H^{\beta}}{T} \quad (2-10)$$

$$\frac{dP}{dT} = \frac{H^{\alpha} - H^{\beta}}{T(V^{\alpha} - V^{\beta})} = \frac{\Delta H}{T\Delta V} \quad (2-11)$$

Eq. (2-11) is the well known Clausius-Clapeyron equation and it is valid for all points along the lines of coexistence. As stated above, it is frequently used as the starting point for vapour pressure correlations. A popular form of Eq. (2-11) is obtained by substituting the compressibility factor for the molar volume, and tidying up the differential on the left hand side:

$$\frac{d \ln P}{d \left(\frac{1}{T} \right)} = - \frac{\Delta H}{R\Delta Z} \quad (2-12)$$

As shown in Figures 2.2 and 2.3 both ΔH_{vap} and ΔZ_{vap} are similar functions of temperature. For this reason the simplest assumption that can be made is that the LHS of Eq. (2-12) is a constant, for the sake of simplicity, called B. Integration of Eq. (2-12) is then trivial:

$$\ln \frac{P}{1 \text{ kPa}} = A - \frac{B}{T} \quad (2-13)$$

This expression can be surprisingly accurate for small enough temperature ranges (typically <20 K, however for certain parts of the temperature range can be as large as 60 K – see Figure 2.4) but for larger temperature ranges it is woefully inaccurate. For this reason various modifications have

been made to increase the accuracy of the predictions. The two most well known of the semi-empirical (Clausius-Clapeyron) type equations are those of Antoine¹⁰ and Riedel¹¹. These two methods and others will be discussed in the sections following.

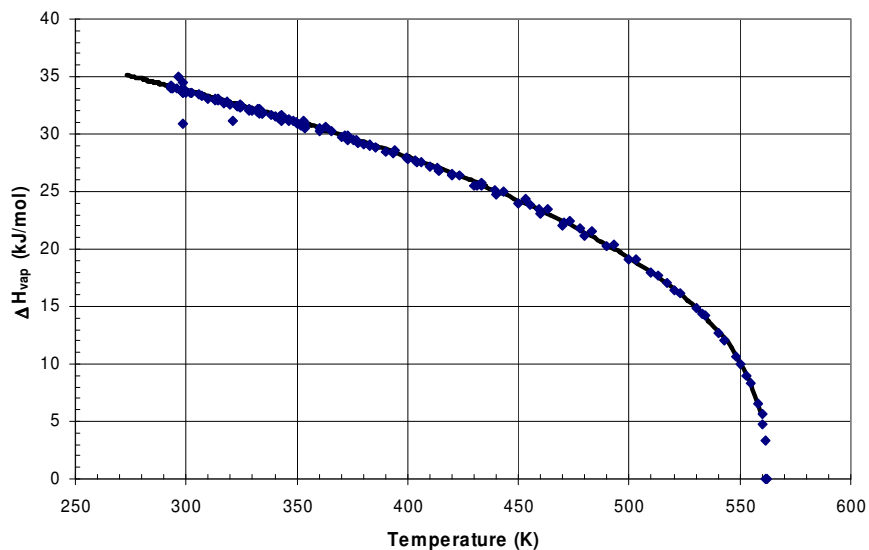


Figure 2.2 Heat of vaporization of benzene as a function of temperature (♦ – data from the DDB², — Watson equation [Eq. (2-36) with $m = 0.391$])

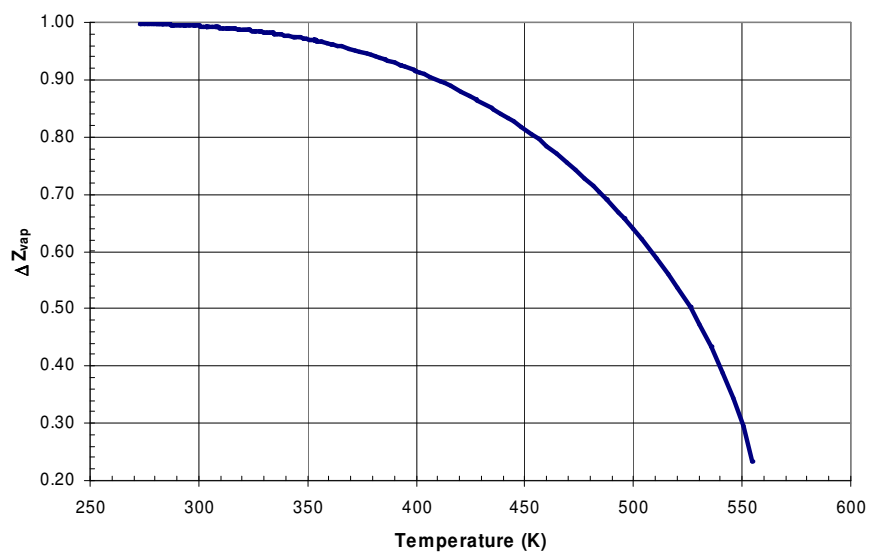


Figure 2.3 ΔZ_{vap} of benzene as a function of temperature using the SRK EOS with Twu alpha function (Twu et al.¹²)

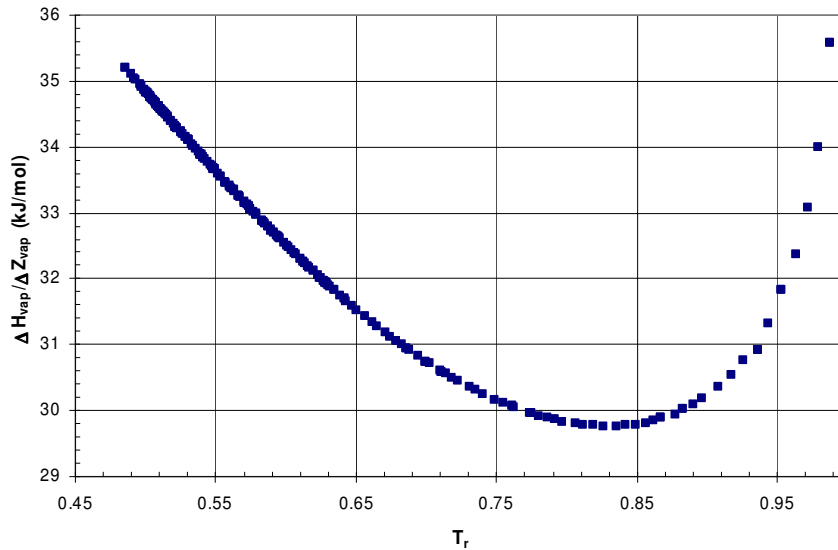


Figure 2.4 $\Delta H_{vap}/\Delta Z_{vap}$ of benzene as a function of reduced temperature ($T_r = T/T_c$) (calculated from the Watson and the SRK EOS using the Twu alpha function (Twu et al.¹²))

Figure 2.4 shows why the assumption of B being a constant is a poor one (except $0.76 < T_r < 0.86$), however is quite evident that the curve is more or less linear below the boiling point ($T_r \approx 0.63$).

2.2.1.1 The Antoine equation

The main problem with Eq. (2-13) is that it is based on assumptions which do not hold. Thus further corrections had to be developed to make the equation more widely applicable. Antoine¹⁰ proposed a new form of Eq. (2-13) as:

$$\log \frac{P^s}{1 \text{ kPa}} = A + \frac{B}{T - C} \quad (2-14)$$

The introduction of the C parameter meant that the equation could now account for the slight bowing of the vapour pressure curve. This equation has since become known as the Antoine equation and has become very popular due to its simplicity and accuracy. The Antoine equation can suffer from poor extrapolation (as with most other fitted models) as shown in Figure 2.5. The equation which was fitted over the full range (270 K – 560 K) of data shows good correlation, even the fairly narrow mid-range data (350 K – 380 K) shows fairly good extrapolation either way. The problem is that when the equation is regressed against either low or high pressure data the extrapolation tends to be fairly poor.

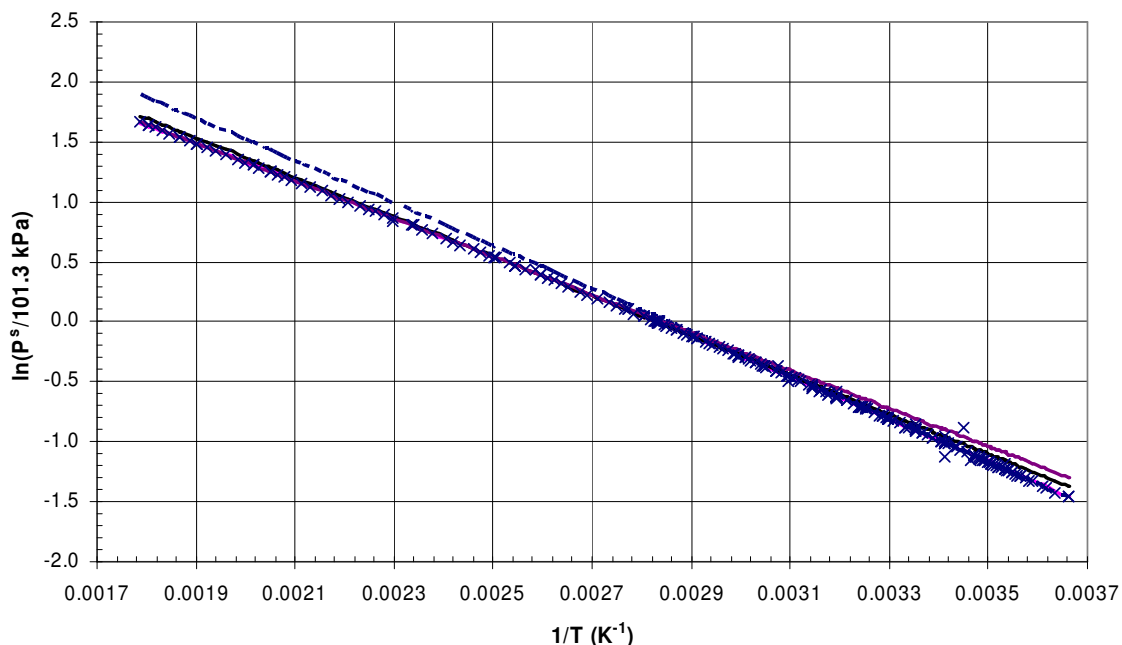


Figure 2.5 $\ln(P^s/101.3 \text{ kPa})$ vs. $1/T$ for benzene showing Antoine plots fitted to different temperature ranges (x – data taken from the DDB², — 270 K to 560 K, - - - 270 K to 300 K, — 350 K to 380 K, — 380 K to 410 K)

Table 2.1 shows how the Antoine constants differ depending on the temperature range used. It is interesting that the C parameter exhibits the greatest instability and for the final two ranges the C gets very small and the Antoine equation approximates Eq.(2-13).

Table 2.1 Antoine constants for various temperature ranges

Range (K)	A	B (K)	C (K)
270 - 560	4.114	-1260.1	47.036
350 - 380	4.957	-1673.3	12.500
270 - 300	4.665	-1647.8	-0.024
380 - 410	4.494	-1580.9	-0.013

Figure 2.6 shows $\Delta H_{vap}/(R\Delta Z_{vap})$ as predicted from the Antoine equation together with the data for benzene. The plot shows the inability of the Antoine equation to predict $\Delta H_{vap}/(R\Delta Z_{vap})$ at high temperatures. Nevertheless the prediction is still fairly accurate for a reasonable temperature above the boiling point. Table 2.2 shows percentage errors of the Antoine equation for various compounds. The Antoine equation compares quite well to more complex methods (in the sections following). For discussion on the errors for this and the other methods presented see Appendix H.

Table 2.2 Relative mean deviations (RMD) for vapour pressures of selected compounds – Antoine equation

Compound Name	NP	RMD (%)
Hexane	282	2.34
Propyl acetate	262	1.68
Benzene	269	1.35
Propanol	282	5.03
Cyclohexane	209	0.87
Perfluorohexane	58	1.94
Methyl isobutyl ketone	84	3.29
1-Chlorohexane	22	1.12

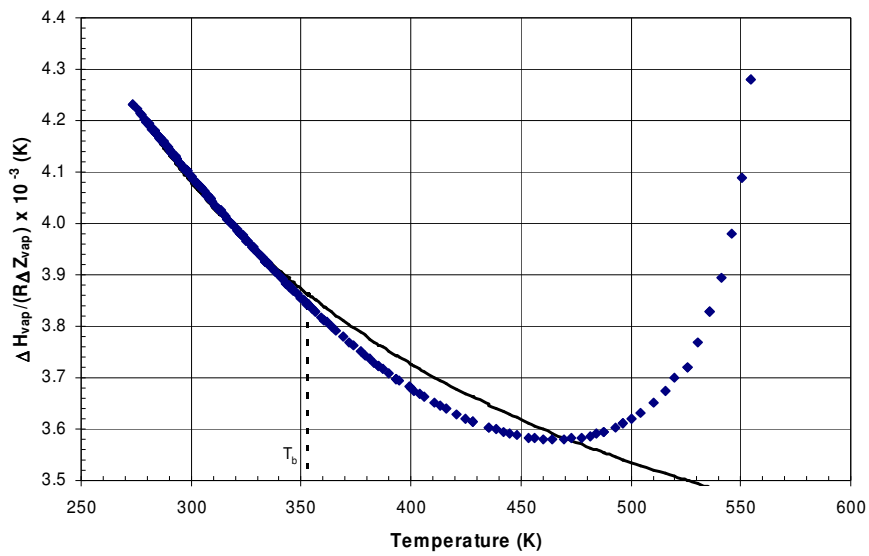


Figure 2.6 Antoine prediction of $\Delta H_{vap}/(R\Delta Z_{vap})$ (♦ - calculated from SRK and the Watson equation for benzene, — Antoine prediction – Eq. (D-1))

2.2.1.2 The Cox equation and Cox charts

The approach taken by Cox¹³ was to rewrite Eq. (2-13) as follows:

$$\log\left(\frac{P^s}{P^{atm}}\right) = A' + \frac{B}{T} \quad (2-15)$$

This means that at the normal boiling point the logarithm will fall away and therefore $B = -A'T_b$, which when substituted back into (2-15) yields:

$$\log\left(\frac{P^s}{P^{atm}}\right) = A' \left(1 - \frac{T_b}{T}\right) \quad (2-16)$$

He then assumed that A' was not a constant, but rather a function of temperature:

$$\log A' = \log A_c + E(1 - T_r)(F - T_r) \quad (2-17)$$

Where A_c is A' at the critical point and E and F are empirical constants. For hydrocarbons with more than two carbon atoms $F = 0.85$. If the critical properties of the substance are not known, a simple power series can be used to approximate A' (the more accuracy required the more terms in the series). For many years the Cox equation was considered to be one of the best equations for vapour pressure for application from the triple point to the boiling point.

Another successful development of Cox¹⁴ was the so called Cox chart. Cox charts are constructed so that, for some reference fluid, the scale of the abscissa is adjusted so that the pressure (log-scale) versus the temperature is a straight line. When other compounds from the same homologous series are plotted the lines are usually found to also be nearly linear. An interesting feature of these plots is that all the lines for a homologous series tend to converge at a point known as the infinite point. Thus for a new compound in the homologous series one only needs a single vapour pressure point to generate an approximate vapour pressure curve. Calingaert and Davis¹⁵ showed that the Cox chart closely represents the Antoine equation. Figure 2.7 shows how the Cox equation provides a more realistic shape of the $\Delta H_{vap}/(R\Delta Z_{vap})$ curve. Even though the shape looks more realistic the error is still comparable to that of the Antoine plot. As can be seen from Table 2.3 the percentage errors are similar to those for the Antoine equation in almost every case.

Table 2.3 Relative mean deviations (RMD) for vapour pressure of selected compounds – Cox equation

Compound Name	NP	RMD (%)
Hexane	282	1.94
Propyl acetate	262	1.91
Benzene	269	1.16
Propanol	282	5.51
Cyclohexane	209	0.68
Perfluorohexane	58	1.81
Methyl isobutyl ketone	84	2.79
1-Chlorohexane	22	0.99

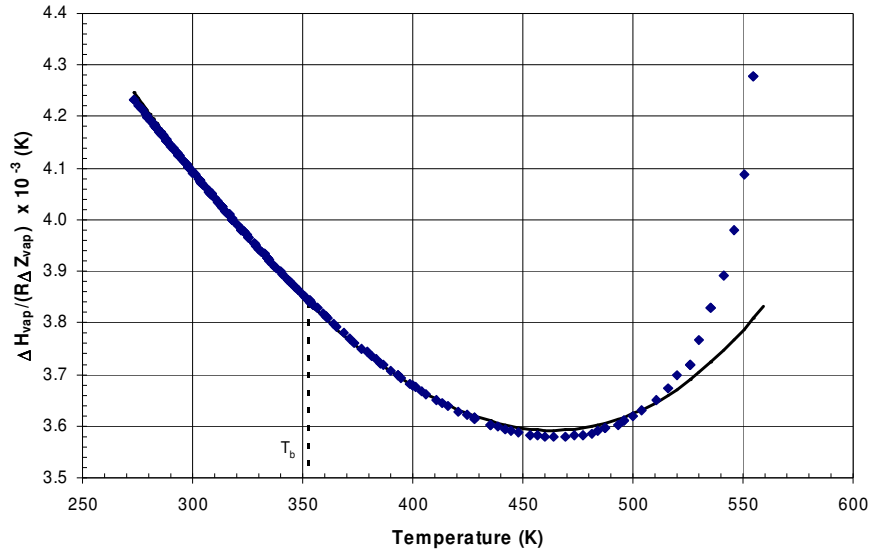


Figure 2.7 Cox prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - calculated from SRK and the Watson equation for benzene, — Cox prediction – Eq. (D-2))

2.2.1.3 The Riedel equation

An approach used by many people is to approximate $B (= \Delta H_{\text{vap}}/R\Delta Z_{\text{vap}})$ with a power series:

$$B = \sum_{i=0}^n B_i T^i \quad (2-18)$$

Integrating Eq. (2-12) using Eq. (2-18) for B gives the following:

$$\ln \frac{P^s}{1 \text{ kPa}} = A + \frac{B_0}{T} + B_1 \ln T + \sum_{i=1}^n \frac{B_{i+1}}{i} T^i \quad (2-19)$$

The widely respected DIPPR¹⁶ group uses a form of Eq. (2-19) known as the DIPPR 101 equation:

$$\ln \frac{P^s}{1 \text{ kPa}} = A + \frac{B}{T} + C \ln T + DT^E \quad (2-20)$$

The Plank-Riedel equation (similar to DIPPR 101. but with a fixed value of E) given in reduced form is:

$$\ln P_r^s = A + \frac{B}{T_r} + C \ln T_r + D T_r^6 \quad (2-21)$$

Based on the Principle of Corresponding States, a criterion known as the Riedel criterion was derived. The Riedel criterion is deduced from plots of α vs. T_r , where α is defined as:

$$\alpha = \frac{d(\ln P_r)}{d(\ln T_r)} \quad (2-22)$$

It states that $\frac{d\alpha}{dT_r} = 0$ when $T_r = 1$ (i.e. at the critical point). Using α_c (which is the value of alpha at the critical point) one can then estimate the values of the Riedel parameters (A,B,C&D). Riedel developed a set of further criteria, which needed to be met in order to obtain a physically realistic vapour pressure equation. (see Appendix C). Figure 2.8 illustrates that the Riedel equation shows a much better reproduction of the experimental shape of the vapour pressure equation than the Antoine equation. The reason that the curve is slightly removed from the data points is that the Riedel equation parameters are calculated from set criteria so as to make the fit physically realistic. Also shown on the plot is the fitted Riedel equation, this was found by using the calculated Riedel parameters as a starting point and regressing the parameters. The resulting curve shows a near perfect representation of the curve up to about 530 K (which is 30 K below the critical temperature). The percentage error for the selected compounds is actually worse than the methods of Antoine and Cox on most occasions. This is not due to the model itself but just the way the parameters are calculated (they are calculated solely from the Riedel criterion and are not fitted to the data).

Table 2.4 Relative mean deviations (RMD) for vapour pressure of selected compounds – Riedel equation

Compound Name	NP	RMD (%)
Hexane	282	2.08
Propyl acetate	262	2.33
Benzene	269	1.99
Propanol	282	7.24
Cyclohexane	209	0.95
Perfluorohexane	58	1.99
Methyl isobutyl ketone	84	4.86
1-Chlorohexane	22	4.48

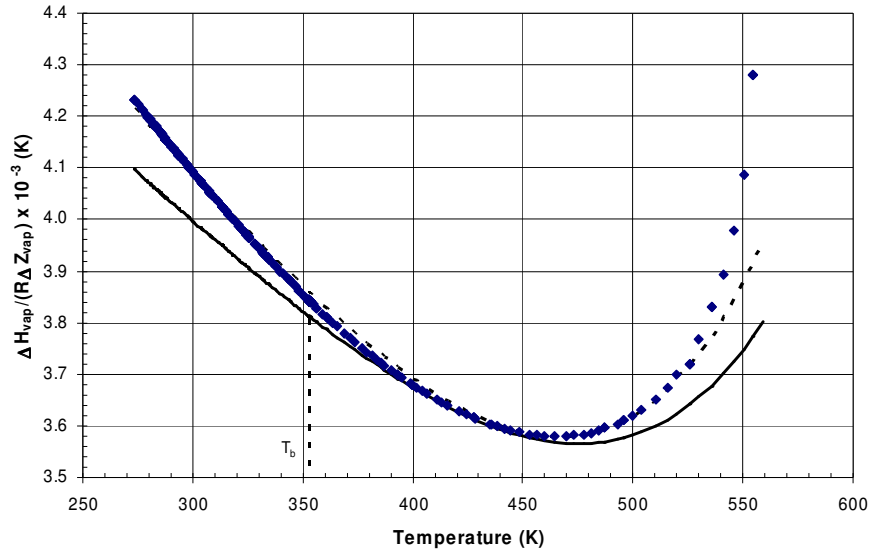


Figure 2.8 Riedel prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Riedel prediction – Eq. (D-3), - - - Riedel direct fit – Eq. (D-3))

2.2.1.4 The Myrdal & Yalkowsky equation

The method of Myrdal & Yalkowsky¹⁷ is a modification of the work of Mishra & Yalkowsky¹⁸. The method is only valid for temperatures below the normal boiling point (the lower the better) since in the model it is assumed that the change in compressibility factor upon vaporization (ΔZ_{vap}) is unity. The form of the Clausius-Clapeyron equation then takes the following form:

$$\ln \frac{P^s}{1 \text{ kPa}} = \int \frac{\Delta H}{RT^2} dT \quad (2-23)$$

The change in enthalpy is then described in terms of the vaporization of a solid (assuming that the heat of vaporization is a linear function of temperature):

$$\Delta H = \Delta H_m + \Delta H_b + (C_p^s - C_p^l)(T_m - T) + (C_p^l - C_p^g)(T_b - T) \quad (2-24)$$

Then assuming that the heat capacities are constant with respect to temperature (which is a reasonable assumption); integrating Eq. (2-23) and introducing the entropy of melting ($\Delta S_m = \Delta H_m / T_m$) and boiling ($\Delta S_b = \Delta H_b / T_b$) the following expression is obtained:

$$\ln \frac{P}{1 \text{ kPa}} = \frac{-\Delta S_m(T_m - T)}{RT} + \frac{\Delta C_{pm}}{R} \left[\frac{T_m - T}{T} - \ln \frac{T_m}{T} \right] - \frac{\Delta S_b(T_b - T)}{RT} + \frac{\Delta C_{pb}}{R} \left[\frac{T_b - T}{T} - \ln \frac{T_b}{T} \right] \quad (2-25)$$

[Note: the derivation shown above is for sublimation, therefore if only vapour and liquid are present the first 2 terms of Eq. (2-25) fall away.] The assumption that is made is that the 4 unknown parameters in Eq. (2-25) can be approximated. The difference between the method of Mydral & Yalkowsky and that of Mishra & Yalkowsky is in the definition of these approximations. Mydral & Yalkowsky introduce some new structural properties like hydrogen bonding number and torsional bond number to more accurately describe the various parameters and make the model more widely applicable. This model has been quite popular in applications involving environmental science. A recent review by Clegg¹⁹ showed that it was comparable to other predictive models such as that of Nannoolal et al.⁸. Eq. (2-25) can be rewritten in the following form:

$$\ln \frac{P}{1 \text{ kPa}} = A + \frac{B}{T} + C \ln T \quad (2-26)$$

where the constants A,B and C are groupings of the parameters in Eq. (2-25). This is to be expected since the heat of vaporization was assumed to be a linear function of temperature (Eq. (2-24)). Therefore the best performance of this model can be calculated by fitting the new equation parameters to data (below the boiling point since that is what the model was designed for). Figure 2.9 shows the best possible Mydral & Yalkowsky prediction for benzene. The parameters were fitted to data below 40 kPa to give the best possible fit. As was expected the fit is good up to approximately the boiling point but diverges greatly thereafter. The method is very simple to implement and provides acceptable errors (Table 2.5) for the selected compounds (only data below the boiling point were used).

Table 2.5 Relative mean deviations (RMD) for vapour pressure of selected compounds - Mydral & Yalkowsky equation

Compound Name	NP	RMD (%)
Hexane	185	13.98
Propyl acetate	151	7.01
Benzene	176	2.73
Propanol	176	11.09
Cyclohexane	158	2.73
Perfluorohexane	35	5.56
Methyl isobutyl ketone	81	3.53
1-Chlorohexane	22	4.85

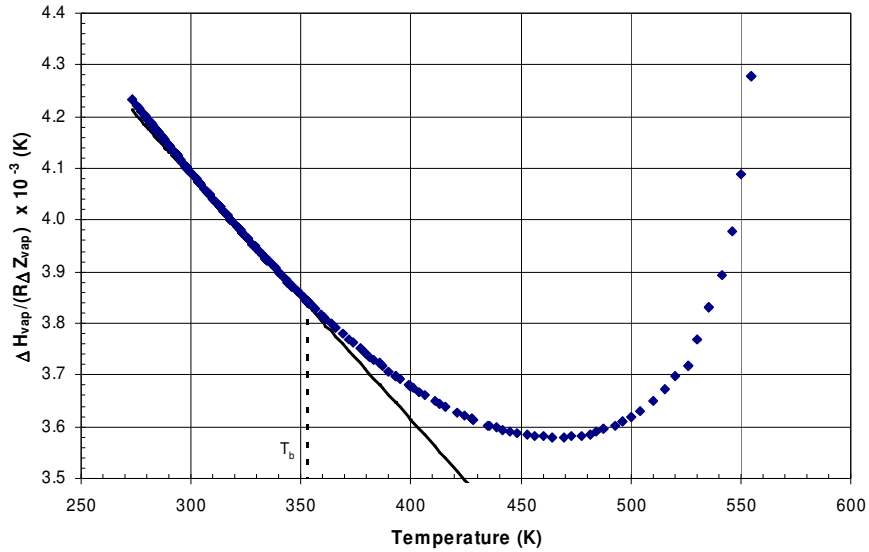


Figure 2.9 Best possible Myrdal & Yalkowsky prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Myrdal & Yalkowsky prediction – Eq. (D-4))

2.2.1.5 The Tu group contribution method

A group contribution method for the estimation of vapour pressures was developed by Tu²⁰. Assuming a quadratic temperature dependence of the B parameter the following vapour pressure equation results (from the integration of the Clausius-Clapeyron):

$$\ln \frac{P^s}{1 \text{ kPa}} = A + \frac{B}{T} - C \ln T - DT \quad (2-27)$$

Then by the usual assumption that the total group contribution is simply the sum of the individual contributions (see paragraph 4.2):

$$\ln \frac{P}{1 \text{ kPa}} = \sum_i N_i \left(A_i + \frac{B_i}{T} - C_i \ln T - D_i T \right) \quad (2-28)$$

It was found, however that this model did not follow the group contribution scheme and therefore the following final equation was used as it followed group contribution much better:

$$\ln\left(\frac{P}{1kPa} \times \frac{M}{g/mol}\right) = \left[\sum_i N_i \left(a_i + \frac{b_i}{T'} - c_i \ln T' - d_i T' \right) \right] + Q \quad (2-29)$$

Where $T' = T/100$, M is the molar mass, a_i , b_i , c_i and d_i are the group contributions for group i ; and Q is a compound specific correction which is given as:

$$Q = \sum_{i=1}^2 \xi_i q_i \quad (2-30)$$

The terms ξ_i and q_i have different functions depending on the value of the index, for $i = 1$ they are structural corrections and for $i = 2$ they are functional group corrections. For alkylbenzenes ξ_1 is given by Eq. (2-31) and for all other compounds $\xi_i = 1$.

$$\xi_1 = s_0 + s_1 N_{cs} + s_2 N_{bs} + s_3 N_{es} \quad (2-31)$$

The N terms are affected by the number and nature of the alkyl substituents. The expressions for q_1 for ring and non-ring compounds respectively are:

$$q_1 = \alpha_{1r} + \frac{\beta_{1r}}{T'} - \gamma_{1r} \ln T' - \delta_{1r} T' \quad (2-32)$$

$$q_1 = \alpha_{1n} + \frac{\beta_{1n}}{T'} - \gamma_{1n} \ln T' - \delta_{1n} T' \quad (2-33)$$

The functional group terms are given as:

$$\xi_2 = f_0 + f_1 N_{cm} + f_2 N_{cm}^3 + f_4 N_{cm}^4 \quad (2-34)$$

$$q_2 = \alpha_2 + \frac{\beta_2}{T'} - \gamma_2 \ln T' - \delta_2 T' \quad (2-35)$$

The term N_{cm} is the total number of carbon atoms in the compound and all the other symbols which have not been explained are simply constants which vary depending on the group of compounds to which they belong. This results in a total of 135 correction parameters and 216 group parameters which means that the total number of model parameters is 351.

The method is claimed to have an average deviation of 5% when tested with 336 organic compounds with 5287 data points. The model parameters were generated by using a set containing 342 compounds with 5359 data points. The high number of model parameters makes this model highly susceptible to over-fitting. Also since many of the parameters are “group-specific” the method becomes less generally applicable. As with the model of Mydral & Yalkowsky the model of Tu was tested by directly fitting Eq. (2-27) to the data and the resulting plot is shown in Figure 2.10. The quadratic approximation increases the capability of the equation up to and just beyond the boiling point but the model still falls off when nearing the critical point. The method of Tu is quite complex to implement by hand and provides very poor prediction for some of the selected compounds.

Table 2.6 Relative mean deviations (RMD) for vapour pressure of selected compounds – Tu equation

Compound Name	NP	RMD (%)
Hexane	282	3.48
Propyl acetate	262	3.39
Benzene	269	19.40
Propanol	282	5.25
Cyclohexane	209	6.86
Perfluorohexane	-	-
Methyl isobutyl ketone	84	14.18
1-Chlorohexane	22	29.82

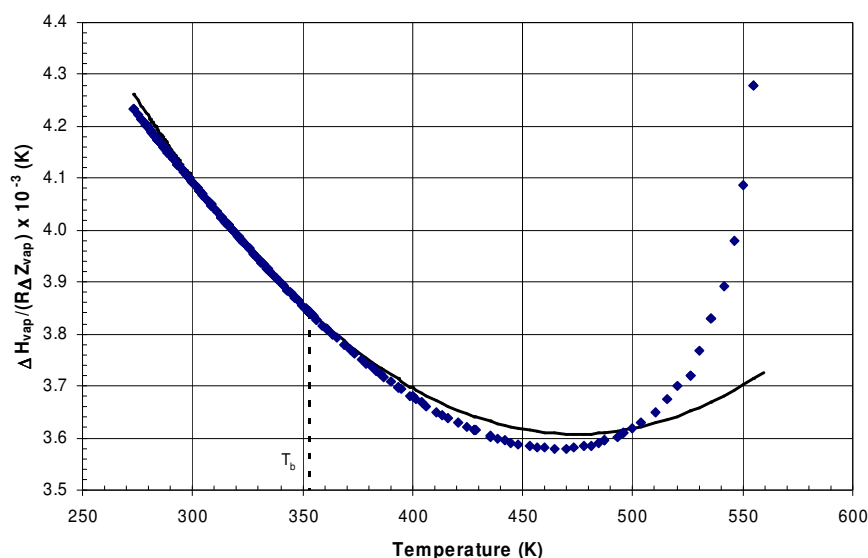


Figure 2.10 Best possible Tu prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Tu prediction – Eq. (D-5))

2.2.1.6 The modified Watson equation

The Watson equation (Eq. (2-36)) is a popular equation for representing the heat of vaporization as a function of temperature. Figure 2.2 shows an example of how accurate the representation can be.

$$\Delta H_v = \Delta H_{vb} \left[\frac{T_c - T}{T_c - T_b} \right]^m \quad (2-36)$$

Lyman et al²¹ assumed that $T_c \approx 1.5T_b$. The Watson equation then simplifies to

$$\Delta H_v \approx \Delta H_{vb} (3 - 2T_{pb})^m \quad (2-37)$$

where $T_{pb} = T / T_b$ and $m = 0.19$. Then combining Eq. (2-37) and Eq. (2-23) (as with Mydral & Yalkowsky this method is only valid for low pressures) integrating twice by parts and dropping the residual integral leads to (the author feels that ΔZ_b may be erroneously included):

$$\ln \frac{P}{P^{atm}} = \frac{\Delta H_{vb}}{\Delta Z_b R T_b} \left[1 - \frac{(3 - 2T_{pb})^m}{T_{pb}} - 2m(3 - 2T_{pb})^{m-1} \ln T_{pb} \right] \quad (2-38)$$

The ΔZ_b term is assumed to always have a value of 0.97, and the following approximation of Fishtine²² is used to calculate $\Delta H_{vb}/T_b$ (where $R = 1.973 \text{ cal/mol.K}$ – as in Eq. (2-38)):

$$\frac{\Delta H_{vb}}{T_b} = K_f (8.75 + R \ln T_b) \quad (2-39)$$

where K_f is dependant on the dipole moment and is tabulated for various compound classes (Lyman et al²¹ and Voutsas et al.²³). As with Mydral & Yalkowsky the advantage of such a method is that only the boiling point is needed to make predictions of the vapour pressure (at low pressures). Figure 2.11 shows how the Watson equation represents the $\Delta H_{vap}/(R\Delta Z_{vap})$ curve. The reason for the rather poor fit is due to the simplifying assumptions that were made in the formulation of the model. Above the boiling point the model falls away drastically and this is due to the fact that an ideal vapour phase (i.e. $\Delta Z = 1$) was assumed. For the results in Table 2.7 only data below the boiling point were used. The errors are fair, however accurate data for K_f is needed and this is difficult for more “exotic” compounds – for example the relatively large error for perfluorohexane.

Table 2.7 Relative mean deviations (RMD) for vapour pressure of selected compounds – Eq. (2-38)

Compound Name	NP	RMD (%)
Hexane	185	4.37
Propyl acetate	151	2.92
Benzene	176	5.89
Propanol	176	8.08
Cyclohexane	158	3.44
Perfluorohexane	35	9.84
Methyl isobutyl ketone	81	2.52
1-Chlorohexane	22	1.89

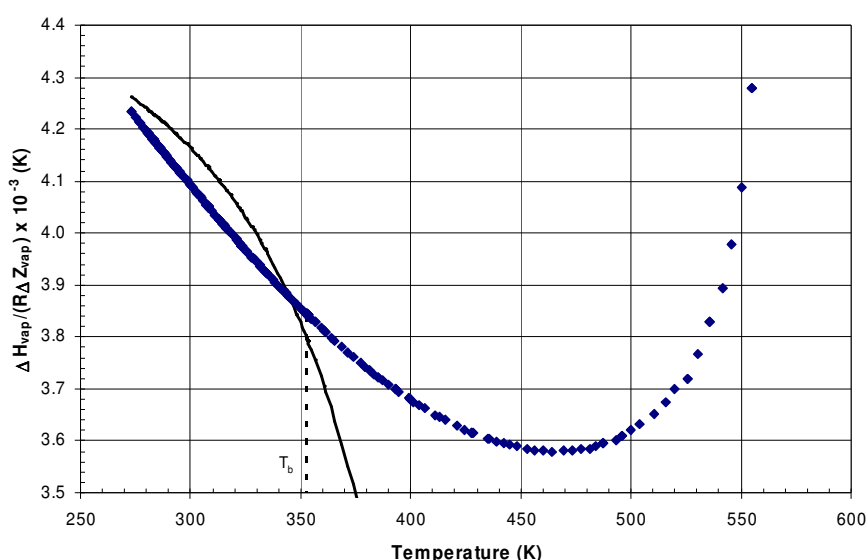


Figure 2.11 Best possible “modified Watson” prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — “Modified Watson” prediction – Eq. (D-6))

2.2.2 Kinetic theory of vaporization

According to kinetic theory, a gas is composed of a large number of molecules which when compared to the distance between them are usually rather small. The molecules are in a constant state of random motion and frequently collide with other molecules and any surrounding objects (i.e. a container wall). The molecules are assumed to have standard physical properties (e.g. mass). The average kinetic energy of the molecules (*viz* velocity) is a measure of the temperature of the gas. Since the particles have mass the collisions with the gas and a surrounding container impart a certain momentum on the container and gives rise to a pressure.

Abrams et al²⁴ developed a vapour pressure equation based on the theoretical treatment of Moelwyn-Hughes²⁵, which uses a multiple-oscillator model for the liquid phase, to take into account for the form of the molecules (using a cubic approximation for B – see Eq. (2-13)):

$$\ln \frac{P}{1 \text{ kPa}} = A + \frac{B}{T} + C \ln T + DT + ET^2 \quad (2-40)$$

The five parameters in the above equation are calculated directly from kinetic theory and so the model has only two adjustable parameters. The first adjustable parameter is s , which is the number of loosely coupled harmonic oscillators in each molecule (this is a model assumption). The second parameter is the characteristic energy E_0 which, together with temperature, is used to measure the rate of molecules escaping into the vapour phase. The expressions for the five parameters are given as follows ($\Gamma(s)$ is the gamma function where $\Gamma(s) = (s-1)!$):

$$A = \ln \left(\frac{R}{V_w} \right) + (s - 0.5) \ln \left(\frac{E_0}{R} \right) - \ln [\Gamma(s)] + \ln \alpha \quad (2-41)$$

$$B = -\frac{E_0}{R} \quad (2-42)$$

$$C = 1.5 - s \quad (2-43)$$

$$D = \frac{s-1}{E_0/R} \quad (2-44)$$

$$E = \frac{2(s-3)(s-1)}{(E_0/R)^2} \quad (2-45)$$

Figure 2.12 shows that this model describes $\Delta H_{vap} / (R \Delta Z_{vap})$ fairly well. The model parameters given by Abrams et al²⁴ were inaccurate and therefore new values of E_0 and s were fitted. The model was only fitted to data below 200 kPa since that is the upper limit of the application of the model. Table 2.8 shows that the model performs very well for the test set of data with most predictions being in the region of 2%. The power of this model is that it has all the benefits of a 5-parameter model (good fit to data) and very few of the downfalls (i.e. stable fits to the data since there are only 2 adjustable parameters – see paragraph 4.1)

Table 2.8 Relative mean deviations (RMD) for vapour pressure of selected compounds – method of Abrams et al.²⁴

Compound Name	NP	RMD (%)
Hexane	203	2.65
Propyl acetate	210	1.95
Benzene	209	1.48
Propanol	273	3.29
Cyclohexane	173	0.83
Perfluorohexane	42	2.16
Methyl isobutyl ketone	85	4.60
1-Chlorohexane	23	2.02

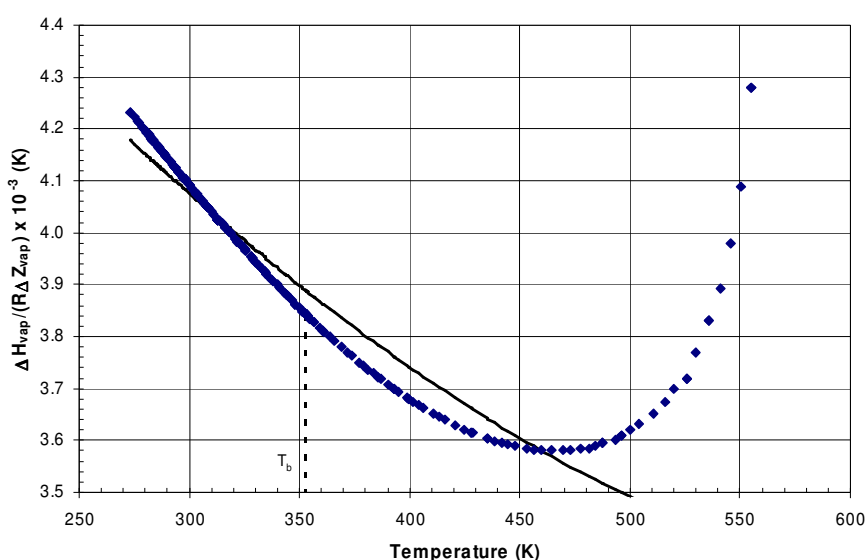


Figure 2.12 Abrams et al.²⁴ prediction of $\Delta H_{\text{vap}} / (R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Abrams et al.²⁴ prediction – Eq. (D-7))

2.2.3 Equations of state

2.2.3.1 Alpha functions

Equations of state are used to relate the macroscopically measurable properties in a system. These properties are usually; temperature, pressure, volume and mass. For a perfect (or ideal) gas the well known Ideal gas law is used. This is the simplest assumption and is only suitable at low pressures and high temperatures. The first cubic equation of state that was applicable to a real gas and the liquid phase was proposed by van der Waals:

$$\left(P + \frac{a}{V_m^2}\right)(V_m - b) = RT \quad (2-46)$$

Where V_m is the molar volume, a is the attraction parameter and b is the repulsion parameter. These parameters are usually calculated from critical temperature and pressure. As cubic equations of state (CEOS) predict identical behaviour for all fluid relative to their T_c and P_c , they are said to employ the 2 parameter corresponding states principle. Since the equation of van der Waals many CEOS have been developed for both pure components and mixtures in the vapour and liquid phase. The most common type of EOS is the cubic EOS, cubic refers to the fact that if expanded the equation would be at most a third order polynomial.

Many of the early equations of state had the downfall that they could not correlate the phase equilibria of mixtures. Soave recognised that the performance for VLE (vapour-liquid-equilibria) was not only dependant on the mixing rule (used to relate pure component parameters to mixture parameters) but also on the performance with respect to pure component vapour pressures. This is accounted for by use of the alpha function in the EOS, an example of such an EOS is the Soave-Redlich-Kwong (SRK) EOS:

$$P = \frac{RT}{V_m - b} - \frac{a\alpha(T_r, \omega)}{V_m(V_m + b)} \quad (2-47)$$

As stated above the alpha function that enabled better vapour pressure representation (for non-polar fluids) was suggested by Soave²⁶ (there were other alpha-type functions that were developed prior to this but they did not provide good enough vapour pressure representation):

$$\alpha = [1 + m(1 - T_r^{0.5})]^2 \quad (2-48)$$

The parameter m is a function of acentric factor as follows (there are 2 versions of this equation):

$$m = 0.480 + 1.574\omega - 0.175\omega^2 \quad (2-49)$$

The Pitzer acentric factor for a pure compound is defined with reference to its vapour pressure. It was noted that $\log P_r$ against $1/T_r$ plots for simple fluids (Ar, Kr, Xe) lay on the same line and passed through the point $\log P_r = -1.0$ at $T_r = 0.7$. The deviation of the non-simple fluids from this point was defined as the acentric factor (Eq. (2-50)), which is sometimes thought of as a measure of the non-sphericity of the molecule (with $\omega = 0$ being perfectly spherical). Equations using the

parameter ω in addition to the T_c and P_c are said to employ the 3 parameter corresponding states principle.

$$\omega \equiv -1.0 - \log(P_r)_{T_r=0.7} \quad (2-50)$$

Soave's alpha function was able reproduce vapour pressure at high reduced temperatures well but diverged for low reduced temperature data (i.e. for the heavy hydrocarbons which have large critical temperatures). There are many different forms of the alpha function that have been proposed since Soave's but, as noted by Coquelet et al.²⁷, they should satisfy the following criterion:

- They must be finite and positive at all temperatures
- They must be assume a value of 1 at the critical point
- The limit as temperature tends to infinity must be zero
- They must be continuous for $T > 0$ as should their first and second derivatives

An example of a further developed alpha function is the 3-parameter Mathias-Copeman²⁸ formulation (notice when $c_2 = c_3 = 0$ it has the same form as Eq. (2-48)):

$$\alpha(T, \omega) = \left[1 + c_1(1 - \sqrt{T_r}) + c_2(1 - \sqrt{T_r})^2 + c_3(1 - \sqrt{T_r})^3 \right]^2 \quad (2-51)$$

The constants (c_1 - c_3) can either be given as values or in the form of generalised equations similar to Eq. (2-49). More recently a very successful alpha function was developed by Twu et al.¹² (Eq. (2-52)). In order to find reliable alpha function parameters that are also valid at very low pressure the regression should include ideal gas and liquid heat capacity at low temperature.

$$\alpha(T) = T_r^{N(M-1)} \exp \left[L(1 - T_r^{NM}) \right] \quad (2-52)$$

Vapour pressures can be found from the EOS by using the fact that $\phi^l = \phi^v$ (since the phases are in equilibrium) and solving the resulting equations for pressure (which will be the vapour pressure). In the absence of a solver there is an iterative routine presented by Reid et al.²⁹. Table 2.9 shows how the Soave EOS performs for the test set. The relatively high error for propanol is due to the fact that this EOS is meant for non-polar molecules. The SRK plot for $\Delta H_{vap} / (R\Delta Z_{vap})$ is shown in Figure 2.13 the prediction is remarkably good below the stationary point as apart from the critical properties and the acentric factor no other data is needed. (The prediction of heat of vaporization is

shown in Appendix F, similar to the method of Eubank et al.^{30,31}).

Table 2.9 Relative mean deviations (RMD) for vapour pressure of selected compounds – SRK EOS using Eqn. (2-48)

Compound Name	NP	RMD (%)
Hexane	282	2.07
Propyl acetate	262	2.51
Benzene	269	1.69
Propanol	282	6.09
Cyclohexane	209	1.47
Perfluorohexane	58	1.79
Methyl isobutyl ketone	84	3.09
1-Chlorohexane	22	1.39

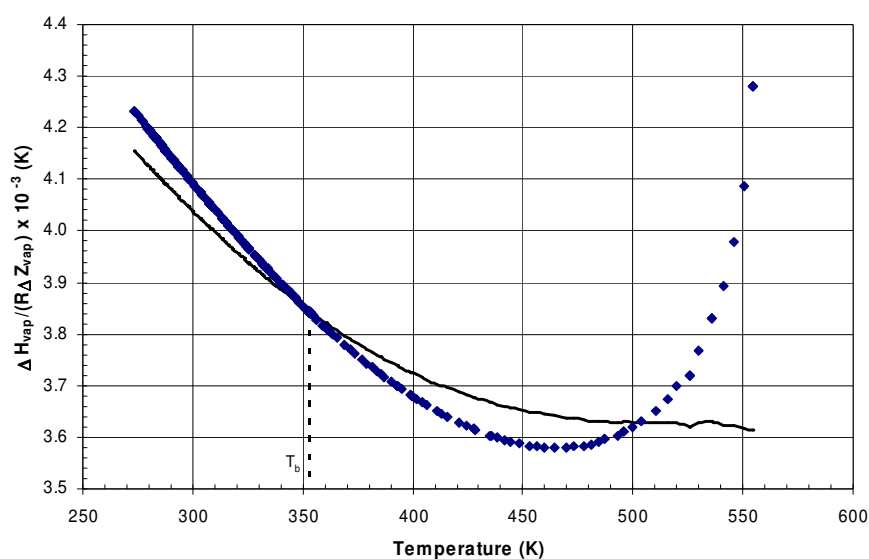


Figure 2.13 SRK prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — SRK prediction – Appendix F)

2.2.3.2 Lee-Kesler method

An alternative to the analytic approach of the cubic EOS is the 3 parameter corresponding states method of Pitzer³², which employs very precise data for two reference fluids. The linear interpolation of Z with respect to 2 reference fluids as a function of the acentric factor is as follows:

$$Z = Z^{(0)}(T_r, P_r) + \omega Z^{(1)}(T_r, P_r) \quad (2-53)$$

The $Z^{(0)}$ term is for simple fluids and the $Z^{(1)}$ term is the departure function for a fluid with $\omega = 1$. The governing assumption is that fluids with the same acentric factor will have the same compressibility factor. Once the compressibility factor is known it can be related to the state variables by its definition:

$$Z = \frac{PV}{nRT} \quad (2-54)$$

The determination of the vapour pressure equation is similar to that of the compressibility factor, with $\ln P_r$ being a linear interpolation between 2 extreme cases:

$$\ln P_r = f^{(0)}(T_r) + \omega f^{(1)}(T_r) \quad (2-55)$$

The functions $f^{(0)}$ and $f^{(1)}$ have are expressed by Lee and Kesler^{33,29} as follows (they have the same form as the Riedel equation):

$$f^{(0)} = 5.92714 - \frac{6.09648}{T_r} - 1.28862 \ln T_r + 0.169347 T_r^6 \quad (2-56)$$

$$f^{(1)} = 15.2518 - \frac{15.6875}{T_r} - 13.4721 \ln T_r + 0.435777 T_r^6 \quad (2-57)$$

As with the alpha functions these equations are popular in the petroleum industry but are limited to fluids for which critical data are available. The Lee-Kesler method was simple to implement and was quite accurate for the set of test data (Table 2.10).

Table 2.10 Relative mean deviations (RMD) for vapour pressure of selected compounds – Lee-Kesler

Compound Name	NP	RMD (%)
Hexane	282	2.49
Propyl acetate	262	2.21
Benzene	269	1.62
Propanol	282	5.98
Cyclohexane	209	2.07
Perfluorohexane	58	2.03
Methyl isobutyl ketone	84	3.39
1-Chlorohexane	22	2.41

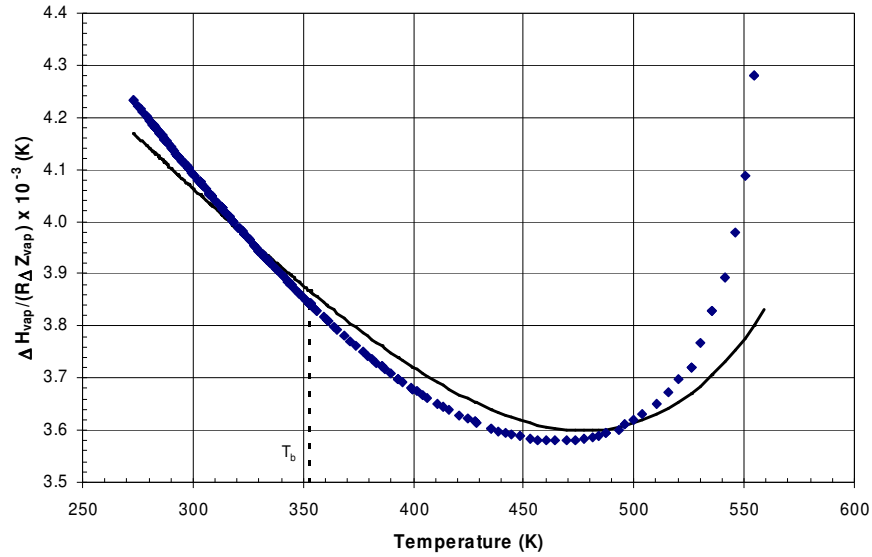


Figure 2.14 Lee-Kesler prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Lee-Kesler prediction – Eq. (D-8))

2.2.4 Empirical models

The very first vapour pressure equation was an empirical correlation suggested by Dalton³⁴ (Eq. (2-58)), however as more accurate vapour measurements were made available his prediction was soon proven to be inaccurate.

$$\ln \frac{P}{1 \text{ kPa}} = A'' + B''T \quad (2-58)$$

Since then many empirical models have been developed for the prediction of the vapour pressure. Another early empirical model is the one developed by Bose³⁴:

$$\ln \frac{P}{1 \text{ kPa}} = a - \frac{b}{T} + \frac{d}{T^2} + \frac{e}{T^3} \quad (2-59)$$

The empirical models that have gained a large amount of respect (and use) are not the ones that are fitted to vapour pressure data indiscriminately but rather those which are subject to certain constraints to make the curve physically realistic.

2.2.4.1 The Wagner equation

Arguably the most popular and widely used empirical method is that proposed by Wagner^{9,35}. Originally the method was developed with very high precision data that was available (i.e. water, nitrogen etc), however since then it has been applied to lower accuracy data. A common form of the Wagner equation is as shown (sometimes referred to as the 1-1.5-3-6 Wagner or just the 3-6 Wagner equation):

$$\ln P_r = \frac{a\tau + b\tau^{1.5} + c\tau^3 + d\tau^6}{1 - \tau} \quad (2-60)$$

with the reversed temperature $\tau = (1 - T_r)$. Fitting of the parameters of the Wagner equation is not simply done by a least squared fit but rather by a more statistical approach where the equation is subject to the following three constraints (as outlined by Chase³⁶):

- The resulting $\Delta H_{\text{vap}}/\Delta Z_{\text{vap}}$ vs. T curve has a minimum within a specified range of reduced temperature values (this is referred to as Waring's³⁷ criterion)
- At low reduced temperatures the value of $\Delta H_{\text{vap}}/\Delta Z_{\text{vap}}$ should approximate ΔH_{vap} (This is because at low reduced temperatures $\Delta Z_{\text{vap}} \rightarrow 1$)
- The term $\ln(P/P')_{T_r=0.95}$ must fall within a specified range, $\ln P'$ is the straight line joining the points $T_r=0.7$ and $T_r=1$ on a $\ln P$ vs $1/T$ plot (This is known as the Ambrose criterion)

The Wagner equation is widely used in industry for the purpose of simulation since it is one of the few equations that can accurately represent the vapour pressure curve from the triple to the critical point. As with the equations of state it can only be applied to compounds for which critical data are available. As previously the validity of the model was analysed by plotting data for $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ against the values predicted from the Wagner equation. The parameters for the Wagner equation were obtained from Reid et al.²⁹. As can be seen in Figure 2.15, the Wagner equation is both accurate and physically realistic. This has a lot to do with the way in which the parameters are fitted (as discussed above). The reason that the curve seems to deviate slightly from the data could be because the Wagner parameters were obtained from an external source and could therefore have been fitted to some lower quality data. Nevertheless it is still quite clear that the Wagner equation provides a physically realistic shape. Even though the Wagner equation has in most cases the lowest error for the test data (Table 2.11), it is still comparable to the other methods that have been presented.

Table 2.11 Relative mean deviations (RMD) for vapour pressure of selected compounds – Wagner (* - parameters fitted by author)

Compound Name	NP	RMD (%)
Hexane	281	2.13
Propyl acetate	262	1.86
Benzene	269	1.15
Propanol	282	3.26
Cyclohexane	209	0.72
Perfluorohexane*	58	2.04
Methyl isobutyl ketone*	84	2.75
1-Chlorohexane*	22	0.89

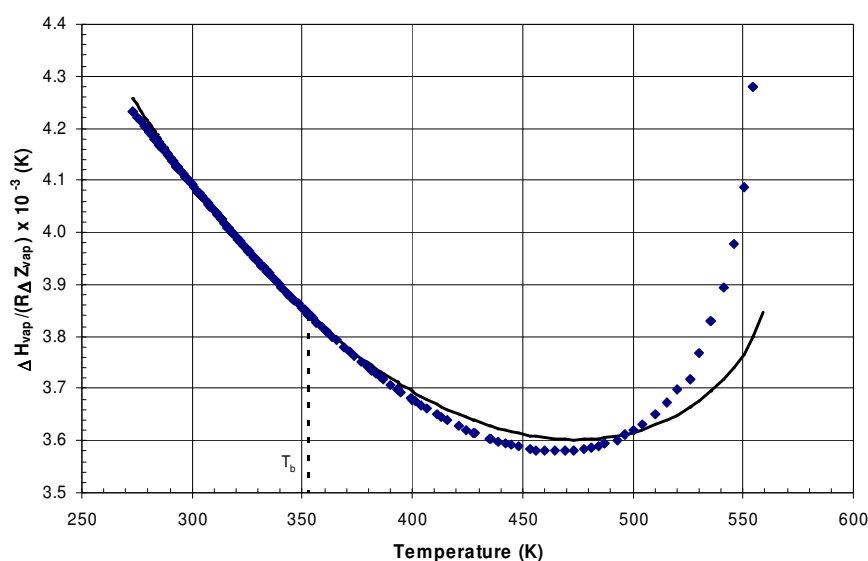


Figure 2.15 Wagner prediction of $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ (♦ - data from SRK and the Watson equation for benzene, — Wagner prediction – Eq. (D-9))

2.2.4.2 Quantitative structure property relationship

Popular methods for property estimation in the environmental and pharmaceutical sciences are quantitative structure property relationships (QSPR). These methods try to relate chemical properties (like vapour pressure and boiling point) to compound specific descriptors like topological indices and polarizabilities. The advantage of using these descriptors is that they can be calculated without any knowledge of the properties of the compound. However as noted by Liang et al.³⁸, some of these methods still rely on some empirical data for example boiling points. Determination of the vapour pressure seems, for the most part, to be restricted to sub-boiling temperatures and in

some cases to a single fixed temperature (usually 25 °C).

A correlation that is proposed by Liang et al.³⁸, relies only on the polarizability and polar functional group counts:

$$\log \frac{P}{1\text{kPa}} = -0.432\alpha - 1.382(OH) - 0.482(C=O) - 0.416(NH) - 2.197(COOH) - 1.383(NO_2) - 1.101(C \equiv N) + 4.610 \quad (2-61)$$

Where α is the polarizability and the bracketed terms refer to the number of the corresponding functional group. The advantage of such an approach is that the vapour pressure can be calculated from knowledge of the chemical structure (since the polarizability can be calculated using quantum mechanical calculations). It is evident from the form of the equation that the vapour pressure is only applicable to a single temperature (25 °C in this case).

Another such model is proposed by Paul³⁹. The parameters in his equation are the solubility parameter (SP), molar refractivity (MolRef), molecular weight (MW) and number of hydrogen bonding acceptors (HBA). Both the solubility parameter and the molar refractivity can be calculated from quantum mechanical calculations. The equation is again for 25 °C and has the following form:

$$\log \frac{P}{1\text{kPa}} = 8.81 + 0.2HBA - 5 \times 10^{-5}(MW)(SP)^2 - 0.05MolRef - 0.08SP \quad (2-62)$$

2.2.4.3 Interpolation polynomials

With the advent of computers, more complex mathematical approaches can be used for the determination of the vapour pressure curve. One of the approaches that are used is to approximate the vapour pressure curve with a very large polynomial series. This is useful because depending on the accuracy needed; terms can be kept or dropped from the series. However this type of approach is quite unstable as one further term can affect all the previously calculated values. Using Chebyshev polynomials can make the equation more stable, this type of approach was outlined by Ambrose⁴⁰.

A novel approach that was found is the use of Interpolation polynomials, where the interpolation polynomial is fixed at certain points and a residual term accounts for any error in the resulting expression. In the equation presented by Ledanois et al.⁴¹ Φ represents the interpolation polynomial

and Γ is the residue term:

$$\ln \frac{P}{1 \text{ kPa}} = \Phi(T_r) + \Gamma(T_r) \quad (2-63)$$

The general form of the interpolation polynomial (in the Newton form) and the residue are:

$$\Phi(T_r) = \sum_{i=0}^{n-1} \frac{\ln P_{i,r}}{\frac{1}{T_{i,r}} - 1} \prod_{j=0}^{i-1} \left(\frac{1}{T_r} - \frac{1}{T_{j,r}} \right) \quad (2-64)$$

$$\Gamma(T_r) = \left[\prod_{j=1}^n \left(\frac{1}{T_r} - \frac{1}{T_{j,r}} \right) \right] \sum_{i=0}^m c_i T_r^i \quad (2-65)$$

The indices n and m in the above equations refer to the number of 'anchor points' and the number of parameters used in the residue function respectively. An approach of this type is useful especially if there are good data available over the full vapour pressure range. It is most often employed together with multi-parameter equations of state, where the solution of the iso-fugacity criterion is numerically difficult. Instead the solutions for the vapour pressure curve are then provided as Chebyshev polynomials for convenience.

2.3 Solvation theory

A method which has gained increasing popularity in recent times is predicting properties from solvation theory. The great advantage of these methods is that they have the potential to be able to produce results for any compound. The exact details of the quantum mechanical calculations are beyond the scope of this work but the general procedure is to calculate the solvation free energy (ΔG^{*sol}) and relate this to the property of interest. Ben-Naim⁴² defines solvation as "The process of transferring one molecule from a fixed position in an ideal gas phase to a fixed position in the fluid phase at constant temperature and pressure". An example of such a model is the one presented by Lin et al.⁴³ where they determine the solvation free energy as follows:

$$\Delta G_{i/i}^{*sol} = \Delta G_{i/i}^{*vdw} + \Delta G_{i/i}^{*el} \quad (2-66)$$

Where the vdw solvation free energy is in effect the non-polar (or van der Waals) contribution to the solvation free energy and the el solvation free energy is made up of the polar and the hydrogen

bonding contributions. Lin et al.⁴³ report a percentage error of 76% for vapour pressures at the normal boiling point which shows that, while promising, these methods still have a long way to go to compete with the current methods available.

2.4 Solid vapour pressures

Many high boiling organic compounds are solid at room temperature (and above). These compounds exert a certain vapour pressure which is of particular interest to environmental scientists who use it to determine the fate of compounds (there are many other uses). Solid vapour pressure data can be converted to a sub cooled liquid vapour pressure which can be represented using the models above. The conversion from solid to sub-cooled liquid data is shown in Figure 2.16. Since the difference between the gradients of the sub-cooled liquid line and the sublimation line is known the conversion is a simple matter. The mathematical representation of this relationship is given by:

$$\ln \frac{P_l}{1 \text{ kPa}} = \ln \frac{P_s}{1 \text{ kPa}} + \frac{\Delta H_m}{R} \left(\frac{1}{T} - \frac{1}{T_m} \right) \quad (2-67)$$

Eq. (2-67) assumes that there are no further transition points (change from one solid phase to another) and that the heat of melting is independent of the temperature which is often an acceptable assumption as the heat capacity changes are relatively small.

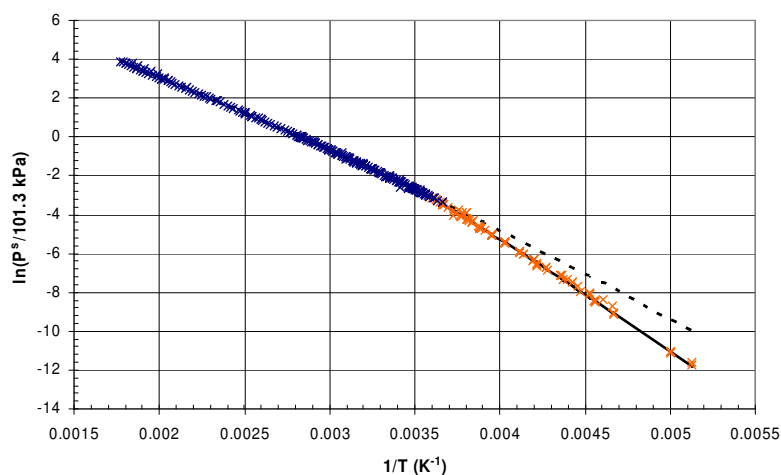


Figure 2.16 $\ln(P^\circ/101.3 \text{ kPa})$ vs. $1/T$ for benzene, with solid vapour pressure data (x – liquid data taken from the DDB²,
 x – solid data taken from the DDB² — solid vapour pressure, - - - sub cooled liquid vapour pressure)

3 COMPUTATION AND DATABASE TOOLS

3.1 Database

The single factor which is most important for the development of a reliable model is the availability of large amounts of accurate data. For this project the Dortmund Data Base (DDB) was utilized. The DDB contains over 180 000 vapour pressure data points (both solid and liquid vapour pressures) for some 6000 compounds.

All data were stored in a Microsoft Access database and Visual Basic for Applications (VBA) in Excel was used to communicate with the database. VBA has many different means to communicate with databases; a popular method is the use of ActiveX Data Objects (ADO). The basic steps that VBA must use when accessing data via ADO are:

- Use ADO to establish a connection to the database
- Create a data storage variable (known as a record set)
- Populate the record set (e.g. via a structured query language (SQL) query)
- Analyse or manipulate the data and save any changes
- Close the record set and terminate the connection

3.2 Data validation

In order to make the model as generally applicable as possible the data that were used for parameter regression needed to be accurate and free of outliers. Due to the massive amount of data it was simply impractical to plot all the data manually and therefore a GUI (graphical user interface) was developed in VBA in order to streamline the whole process. Plotting data is a very good way of removing any obvious outliers, however due to the logarithmic scale high pressure outliers were more difficult to notice. While plotting of the data is a very practical and fast way of detecting outliers there is sometimes simply no way to determine the accuracy of the data. The reason for this is that if there is only data from one source, the data may be smooth and therefore seem accurate but may in fact have been measured incorrectly. (Screenshots of the data validation GUI and the model testing GUI are shown in Appendix E)

Consider the following two typical examples of data found in the DDB. Figure 3.1 shows the plot of the data for amyl formate. In the absence of any accurate boiling point data one may consider the upper group of data to be the more accurate (as it contains the majority of the data points). Fortunately in this example, the quality assessment in the DDB has marked the above data as

questionable and therefore they were removed from the training set. Typically this type of error could be resolved by either the internal quality assessment of the DDB or the availability of accurate boiling point data. Figure 3.2 shows the plot of the data for n-eicosane (C-20 alkane). There is a large amount of scatter in the data with no real way to distinguish between good and poor data (apart from some of the obvious outliers). For cases such as these data was either removed (for use in the test dataset) or if the scatter was not too severe and the compound was “fairly exotic” (i.e. contained unusual functional groups) it was included in the training set.

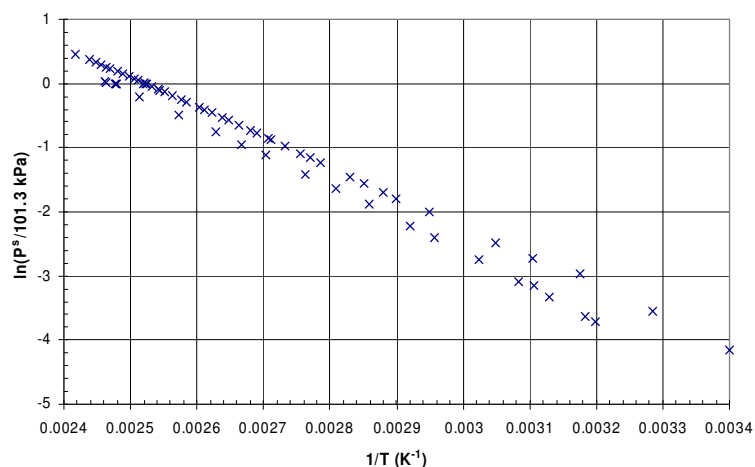


Figure 3.1 Experimental data from the DDB² for amyl formate

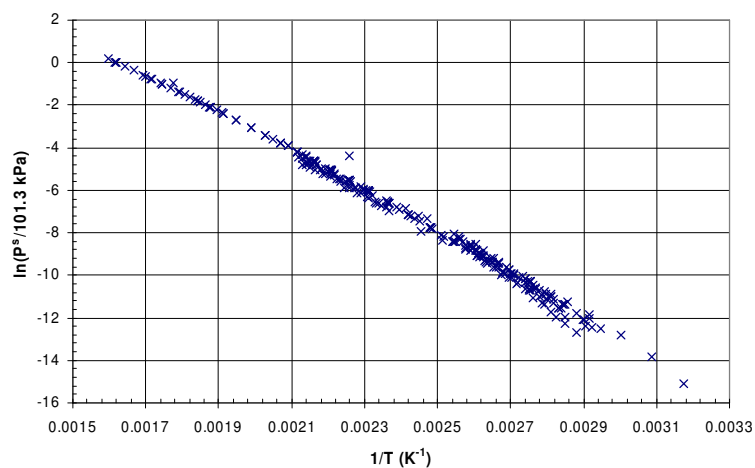


Figure 3.2 Experimental data from the DDB² for n-eicosane

3.3 Regression

Since this project essentially entails model development, it is obvious that there is a certain amount of “curve-fitting” required. The regression techniques described in the sections following are by no means the only solution to parameter evaluation but are the ones that were thought to be most suitable to the task. Both the linear and the non-linear routines were taken from the book by Press et al.⁴⁴ (only the algorithms and not the code).

3.3.1 Linear regression

The simplest case of regression is linear regression, whereby the model must be linear in the parameters to be regressed as follows:

$$y(x) = \sum_{k=1}^M a_k X_k(x) \quad (3-1)$$

The model does not have to be linear in x , since $X(x)$ can be any function of x , for example the following expression could result:

$$y(x) = a_0 + \frac{a_1}{x} + a_2 \ln x \quad (3-2)$$

The next step is to define an objective function (sometimes called a merit function, F) which is a measure of the deviation of the fitted model from the experimental data (y_i):

$$F = \sum_{i=1}^N \left[\frac{y_i - \sum_{k=1}^M a_k X_k(x_i)}{\sigma_i} \right]^2 \quad (3-3)$$

The term σ_i is the standard deviation of the i^{th} data point. From this point onwards it will be disregarded ($\sigma_i=1$) since it was not used in this work.

We then define an $N \times M$ matrix **A**, and a vector **b** of length N as follows (N is number of data points and M is number of parameters; for any meaningful results $N > M$):

$$A_{ij} = X_j(x_i) \quad (3-4)$$

$$b_i = y_i \quad (3-5)$$

The “best-fit” occurs when the derivative of the objective function, with respect to the regression parameters, falls to zero, therefore:

$$\frac{\partial F}{\partial a_k} = -2 \sum_{i=1}^N X_k(x_i) \left[y_i - \sum_{j=1}^M a_j X_j(x_i) \right] = 0 \quad (3-6)$$

Interchanging the order of the summations, simplifying and rearranging yields:

$$\sum_{j=1}^M \alpha_{kj} a_j = \beta_k \quad (3-7)$$

where the expressions for α and β , in matrix form, are as follows:

$$[\alpha] = \mathbf{A}^T \cdot \mathbf{A} \quad (3-8)$$

$$\{\beta\} = \mathbf{A}^T \cdot \mathbf{b} \quad (3-9)$$

Rewriting Eq. (3-7) in matrix form we get,

$$[\alpha] \cdot \mathbf{a} = \{\beta\} \quad (3-10)$$

which can be solved for \mathbf{a} by, for example, Gauss-Jordan elimination

$$\mathbf{a} = [\alpha]^{-1} \{\beta\} \quad (3-11)$$

3.3.2 Non-linear regression

For systems where the model is non-linear with respect to its parameters the method of linear least squares does not apply, and more complex algorithms must be used. Unlike the linear least squared routine there can be (and very often are) multiple minima for the objective function, and therefore the solution which the routine converges upon may only be a local minimum as

opposed to the global minimum. Many such algorithms exist but there is no one algorithm which has been shown to be completely suitable in all circumstances. One of the most popular non-linear routines is the one developed by Levenberg and Marquardt. The routine is widely reputed to be robust and globally convergent. Globally convergent refers to the fact that the algorithm can converge on a minimum from anywhere on the domain (it unfortunately does not mean that it converges on the global minimum).

Since the model to be fitted is non-linear in its parameters the objective is defined in general form as:

$$F(\mathbf{a}) = \sum_{i=1}^N [y_i - y(x_i; \mathbf{a})]^2 \quad (3-12)$$

When \mathbf{a} is close enough it is expected that the objective function will be well approximated by the following quadratic (\mathbf{D} is $M \times M$ and \mathbf{d} has M elements):

$$F(\mathbf{a}) \approx \gamma - \mathbf{d} \cdot \mathbf{a} + \frac{1}{2} \mathbf{a} \cdot \mathbf{D} \cdot \mathbf{a} \quad (3-13)$$

As with all “Newton-type” methods the solution is found by taking a step down the path of steepest decent, which can be written as:

$$\delta \mathbf{a} = \mathbf{D}^{-1} \times [-\nabla F(\mathbf{a}_{cur})] \quad (3-14)$$

When the quadratic approximation is good then the function will converge in one step. \mathbf{D} is known as the Hessian matrix and can be found from the second derivative of the objective function. The first derivative of the objective function (which must disappear at the minimum) is given as:

$$\frac{\partial F}{\partial a_k} = -2 \sum_{i=1}^N [y_i - y(x_i; \mathbf{a})] \frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \quad k = 1, 2, \dots, M \quad (3-15)$$

Taking one further partial derivative results in the Hessian matrix:

$$\frac{\partial^2 F}{\partial a_k \partial a_l} = 2 \sum_{i=1}^N \left[\frac{\partial y(x_i; \mathbf{a})}{\partial a_k} \frac{\partial y(x_i; \mathbf{a})}{\partial a_l} - [y_i - y(x_i; \mathbf{a})] \frac{\partial^2 y(x_i; \mathbf{a})}{\partial a_k \partial a_l} \right] \quad (3-16)$$

This equation can become very expensive (in terms of computing time) when the function is complex and the number of variables is high. For this reason the second derivative term is dropped. Some authors⁴⁵ suggest that it makes it more robust however, for this work, no advantage was found. As can be seen from Eq. (3-14) the effect of the Hessian is to determine the step size and therefore dropping the double derivative term will only affect the path taken and not the final solution. To make the routine as general as possible (to enable the testing of multiple models) a numerical approximation was used for the partial derivatives of the model. A simple “backwards difference” approximation Eq. (3-17) was made for the first derivative since this results in the fewest function calls and as stated above the Hessian need not be perfectly accurate.

$$\frac{\delta y}{\delta x} = \frac{y_m - y_{m-1}}{\Delta x} \quad (3-17)$$

To avoid the rather clumsy derivative notation 2 parameters α and β are defined (not the same as those defined above):

$$\alpha_{kl} \equiv \frac{1}{2} \frac{\partial^2 F}{\partial a_k \partial a_l} \quad (3-18)$$

$$\beta_k \equiv -\frac{1}{2} \frac{\partial F}{\partial a_k} \quad (3-19)$$

The modification of Levenberg-Marquardt was the introduction of the procedure of damping and boosting the step size of the minimum search. This is done by altering α by defining a new term α' as (other boosting and damping schemes are presented by Lampton⁴⁶):

$$\begin{aligned} \alpha'_{jj} &\equiv \alpha_{jj}(1 + \lambda) \\ \alpha'_{jk} &\equiv \alpha_{jk} \quad (j \neq k) \end{aligned} \quad (3-20)$$

The steepest decent formula Eq. (3-14) may then be written in terms of α' and β as:

$$\delta \mathbf{a} = [\alpha']^{-1} \cdot \{\beta\} \quad (3-21)$$

The general procedure for the Levenberg-Marquardt routine is as follows:

1. Chose an initial value for λ (for example $\lambda = 0.00001$)

2. Set meaningful starting values for the parameters in \mathbf{a}
3. Calculate the value of the objective function
4. Evaluate $\delta\mathbf{a}$ by Eq. (3-21) and calculate $F(\mathbf{a} + \delta\mathbf{a})$
5. If $F(\mathbf{a} + \delta\mathbf{a}) \geq F(\mathbf{a})$ then increase λ by some factor (10 is popular) and return to step 3
6. If $F(\mathbf{a} + \delta\mathbf{a}) \leq F(\mathbf{a})$ then decrease λ by some factor (again 10 is popular)
7. Set $\mathbf{a} = \mathbf{a} + \delta\mathbf{a}$ and check some convergence criterion (e.g. $\delta\mathbf{a} \rightarrow 0$) if the solution has adequately converged terminate the loop otherwise return to step 3

3.3.3 Inside-Outside regression

In some cases there are a large number of model parameters which are linear and only a few which are not. This then means that a slow (compared to linear regression) non-linear regression must be used to evaluate all the parameters. One way to circumvent this problem is to use a combination of the two algorithms in a so called "Inside-Outside" type regression. The term refers to the fact that there are 2 (or more) nested loops in the procedure with the outside loop only running once the inside loop(s) has converged. A flow diagram of the general procedure is shown in Figure 3.3. This procedure can shorten the regression time from a matter of hours (in some cases even days) to a matter of minutes. The reason for this is that the (expensive) LM regression is only used for the few non-linear parameters while the fast linear least squares is used for the rest.

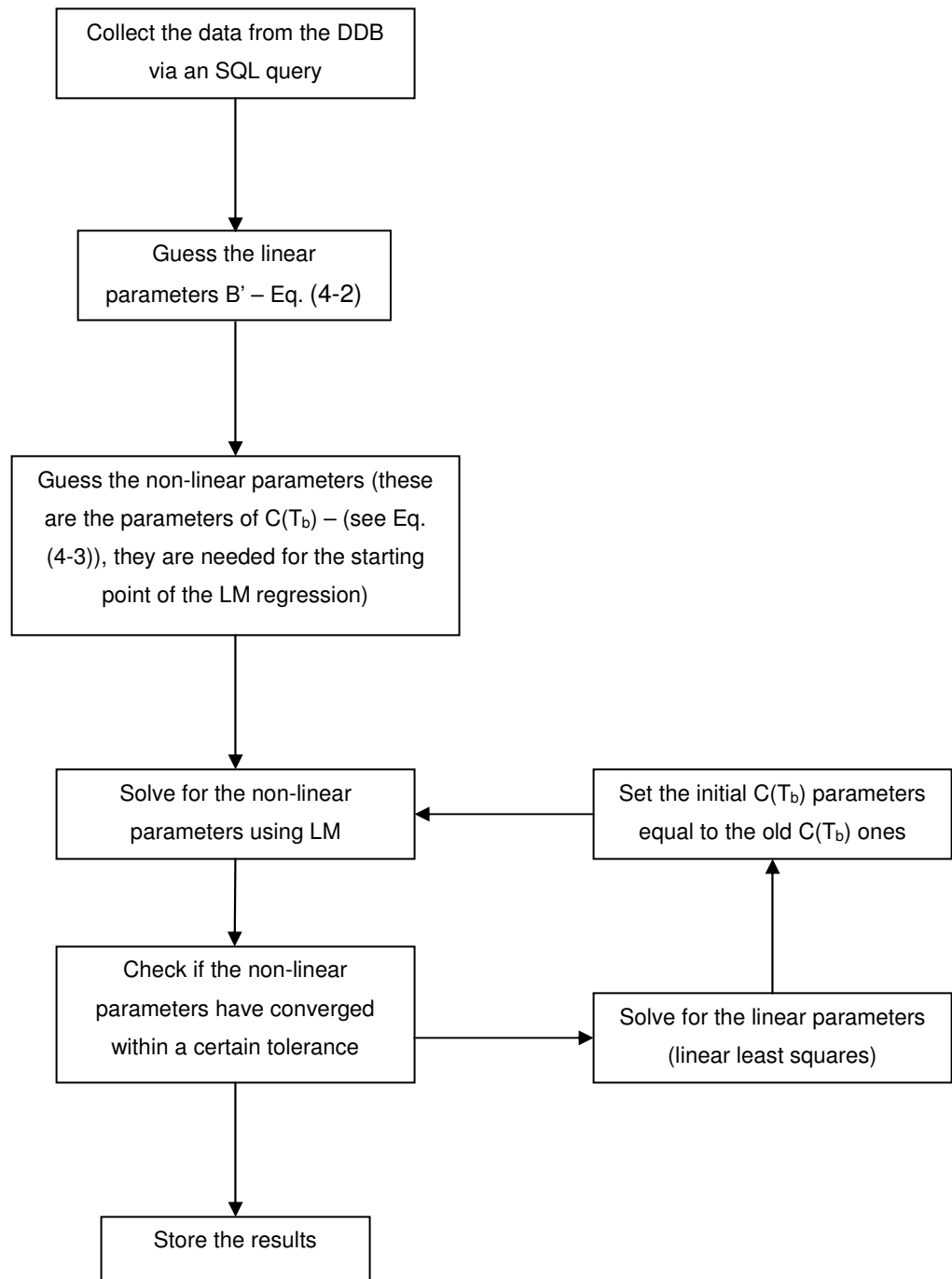


Figure 3.3 Flow diagram for the “Inside-Outside” regression technique

3.3.4 Implementation

Both the non-linear and linear least squared fits were coded in Compaq Visual Fortran (CVF), since Fortran is well known to be able to handle computations very efficiently. CVF did not provide a simple way to access databases so since all the other data processing was done in VBA, the

regression vectors and matrices were formed in VBA and passed to a Fortran dll (a dynamic link library). Since there are huge amounts of data being passed to the dll many of the arrays used needed to be allocatable (only allocate memory when needed).

3.3.5 Fragmentation

Group contribution is based on the assumption that molecules can be broken down into groups which can be used to describe the behaviour of the molecule. Therefore in order to develop a group contribution method for any reasonable number of groups, automatic fragmentation software is required. DDBST⁴⁷ had developed a software package which can carry out the fragmentation into functional groups. The number and type of groups can be manipulated by changing an “ink-file” (called this because of the German word *inkrement*). When changing the ink-file two considerations should be made; firstly the priority of the group and secondly the group definition. The group priority determines which group is fragmented first, for example a COOH group would be fragmented before an OH or a ketone group otherwise the program will never find any COOH groups. The group definition is also very important and must follow a strict format as shown for the example of an aliphatic carboxylic acid group:

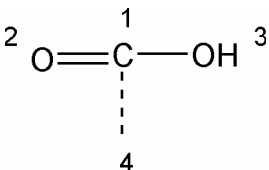
(1) Aliphatic Carboxylic Acid §COOH§	
(2) 4 3 53 53	
(3) C 3 2 K 0 Ja	
(4) O 1 1 K 0 Ja	
(5) O 1 1 K 0 Ja	
(6) C 4 1 N 0 Nein	
(7) 1 2 2 K	
(8) 1 3 1 K	
(9) 1 4 1 K	

Figure 3.4 The group definition and structure of the aliphatic carboxylic acid group

Line 2: General description of the group, the first number is the number of atoms in the group, second number is the number of bonds, the third and fourth are the group number (always the same for this method but different, for example, in the case of UNIFAC which employs sub and main groups)

Line 3 – 6: Description of the group atoms, first character is the atom symbol; second is the maximum number of substituents; third is the minimum number of substituents; forth is the type of atom K,N,R or A are used to represent chain, non-aromatic, ring or aromatic respectively (a * is used when the type of atom is inconsequential); fifth is the charge which is 0 for all groups in this method and sixth is *Ja*(yes) or *Nein*(no) to determine whether to include the atom in the group or

not (in this example the fourth atom is not included since it is not part of the group but still important for the correct definition of the group)

Line 7 – 9: Description of the bonds in the group, the first 2 numbers are the atom numbers between which the bond occurs; the third number is the number of bonds (i.e. single, double etc); the last character is the type of bond (A, R or K used – same meaning as previously).

4 DEVELOPMENT OF THE METHOD

4.1 Model development

It is fairly intuitive that the accuracy of the regression should be increased with the number of model parameters. However the more parameters there are the more difficult it becomes to accurately fit the model parameters. The reason for this is that very often the parameters are intercorrelated, i.e. they contribute to the same effect in some way. This means that there are some parameter values that produce very good fits but are physically very unrealistic. This became very apparent when fitting the equations in Section 2.2; consider the example of the model given by Eq. (2-27). Figure 4.1 and Figure 4.2 show the marked difference in the plots for $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$, however the vapour pressure plots look remarkably similar. Even though the vapour pressure fit in Figure 4.1 is slightly better than that of Figure 4.2 it shows that seemingly correct parameters can be very wrong. This problem is made much worse when fitting to only a small section of the data.

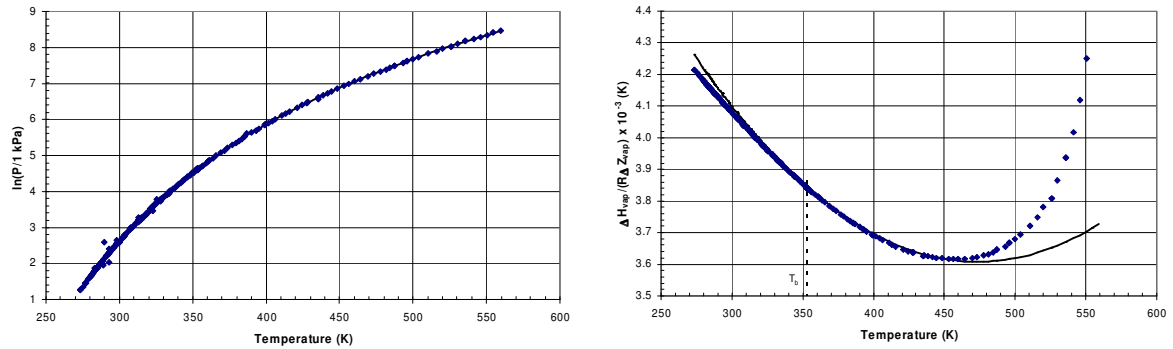


Figure 4.1 A proper fit for the Eq. (2-27)

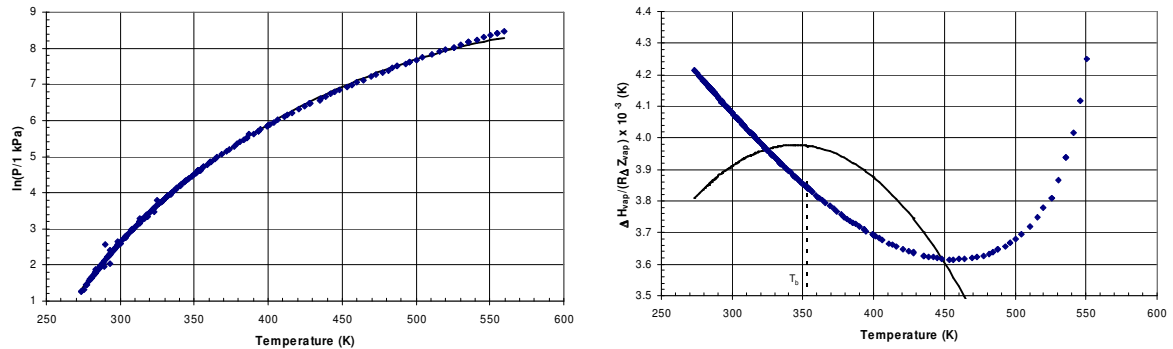


Figure 4.2 A physically unrealistic fit for Eq. (2-27)

It is for this reason that the objective of this work was to develop an equation with as few parameters as possible. From all the equations presented above the one which stands out for both its simplicity and accuracy is the Antoine equation. At very low temperatures (in the vicinity of $T=C$) the Antoine equation diverges, however since C is typically in the range of 40K to 70K this is outside the range of practical interest (but may still cause problems, e.g. in simulation iterations). The problem of intercorrelated parameters is still a weakness, and therefore the Antoine equation needed to be modified. An obvious choice would be to use the normal boiling point instead of the parameter A since there are a large number of normal boiling point data available in literature. This is done by using the normal boiling point as a datum point, this results in the following equation:

$$\ln\left(\frac{P}{P_{atm}}\right) = \frac{B}{T-C} - \frac{B}{T_b-C} = \frac{-B}{T_b-C} \times \frac{T-T_b}{T-C} \quad (4-1)$$

This new model however still suffers from the fact that the model parameters B and C need to be regressed to data (and since B and C are intercorrelated no meaningful group contribution method can be developed). It has been observed by Thomson³⁴ that the C parameter correlates with the normal boiling point and the equation can be written in the following form:

$$\ln\left(\frac{P}{P_{atm}}\right) = B' \frac{T-T_b}{T-C(T_b)} \quad (4-2)$$

While the value of C was simply assumed to be $T_b/8$ in the model of Nannoolal et al.⁸, the following function was found to give better representation of both large and small molecules as well as providing greatly improved representation of low pressure data:

$$C(T_b) = -2.65 K + \frac{T_b^{1.485}}{135 K^{0.485}} \quad (4-3)$$

The advantage of this C -parameter correlation is that it not only provides a better representation of the data but also improves the group contribution estimation of B' . As with the model of Nannoolal et al.⁸, Eq. (4-2) cannot model the aliphatic alcohols and carboxylic acids correctly and for this reason a logarithmic correction term was added (for all other compounds D' is set to zero):

$$\ln\left(\frac{P}{P_{atm}}\right) = B' \frac{T-T_b}{T-C(T_b)} + D' \ln\left(\frac{T}{T_b}\right) \quad (4-4)$$

The effect of this correction term is very significant and therefore the predictions of the aliphatic alcohols and carboxylic acids are dramatically improved.

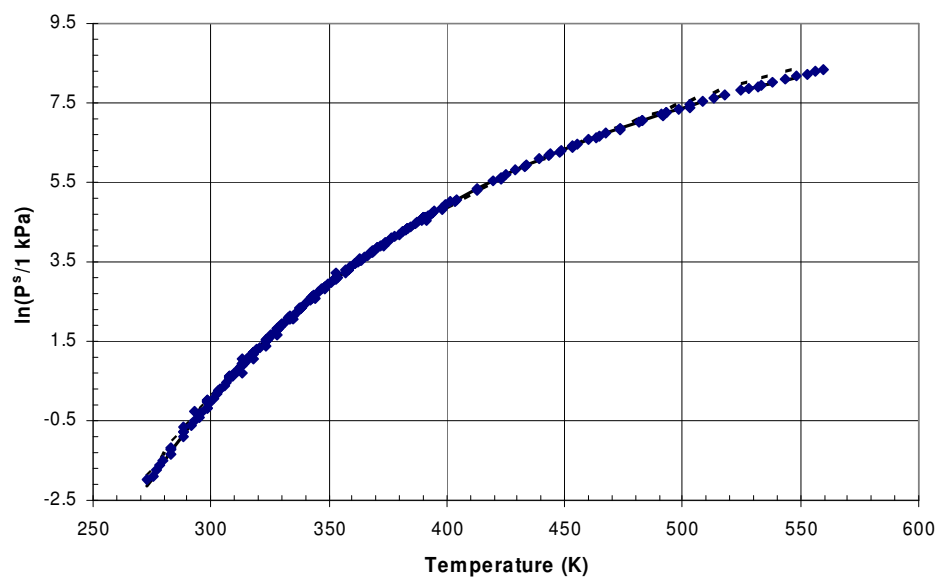


Figure 4.3 $\ln(P^s/1 \text{ kPa})$ vs. T for 1-butanol (♦ - data from the DDB, — with the logarithm term, - - - without the logarithm term)

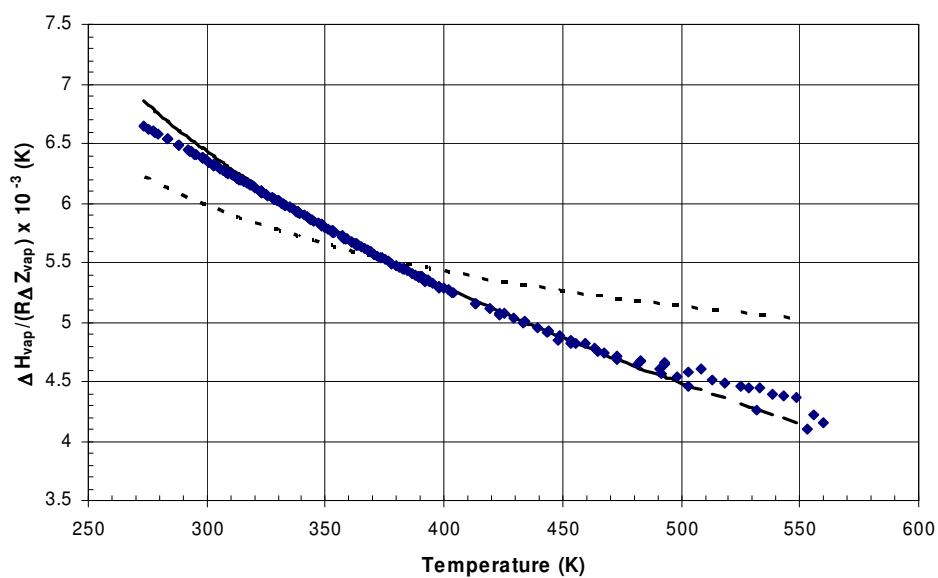


Figure 4.4 $\Delta H_{\text{vap}}/(R\Delta Z_{\text{vap}})$ for 1-butanol (♦ - data from SRK using the MC^{28} alpha function and the Watson equation (Eq. (2-36) with $m = 0.473$), — prediction with the logarithm term, - - - prediction without the logarithm term)

Consider the example of 1-butanol; Figure 4.3 shows the improved vapour pressure fit (the error is reduced from 7.5% to 2.8%). This improvement is not so visible from the vapour pressure plot but the plot of $\Delta H_{vap}/(R\Delta Z_{vap})$ quite clearly shows the substantial improvement that the logarithmic term makes to the physical realism of the model parameters (It is interesting to note that even a direct fit of all three Antoine parameters does not give an adequate description of the $\Delta H_{vap}/(R\Delta Z_{vap})$ curve and therefore a logarithmic correction term is also needed; in this form it is known as the modified Antoine.). Both B' and D' in Eq. (4-4) are calculated from group contribution. Due to significant intercorrelation simultaneous regression of B' and the group increments for D' was required.

In an attempt to make the model more widely applicable an effort was made to correlate the model parameters with properties which can be predicted *ab initio*. One such property is the polarizability, which is basically the tendency of a molecule to be polarized by an external electric field. The polarizability data was taken from *ab initio* DFT calculations using the hybrid functional B3LYP and the electron representation 663B in the program Gaussian 2003. As shown in Figure 4.5 the correlation is very good for the n-alkanes, however when the data for the hydrocarbons is plotted (Figure 4.6) it is clear that one can not draw any meaningful correlations from the data. There was a fairly good correlation between the boiling point and polarizability for the hydrocarbons as shown in Figure 4.7, however no meaningful results could be obtained as the scatter was too high.

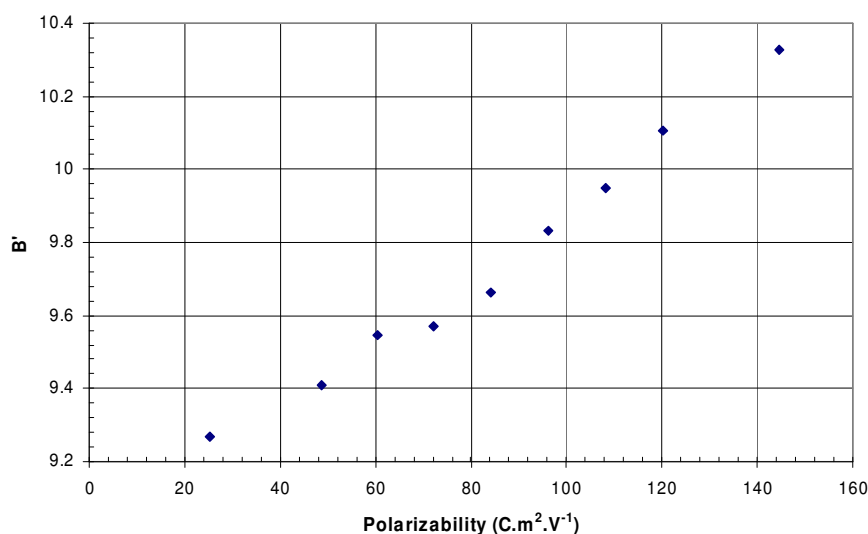


Figure 4.5 B' vs. polarizability for the n-alkanes

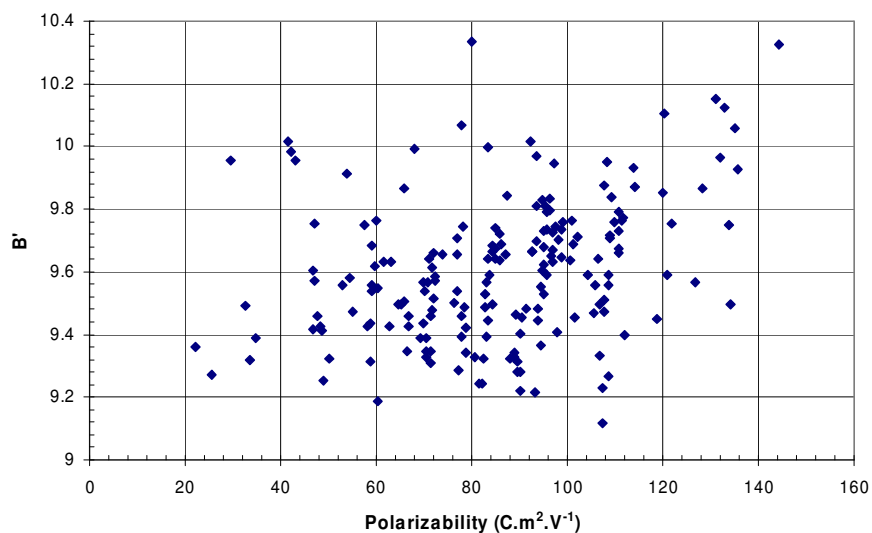


Figure 4.6 B' vs. polarizability for hydrocarbons

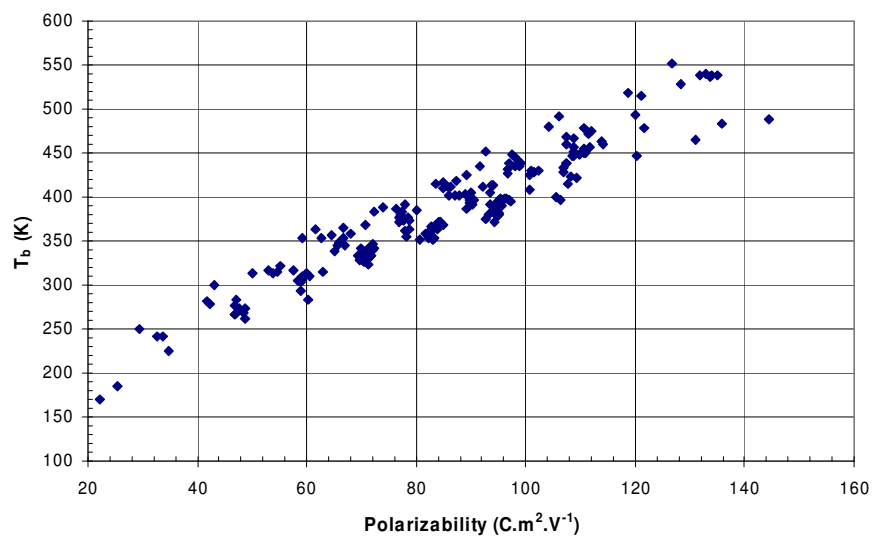


Figure 4.7 T_b vs. polarizability for hydrocarbons

Beside the advantage of having the parameters predicted with no experimental data, an approach like this also would enable the model to be split into different part for different effects (non-polar, polar and hydrogen bonding). The advantage of this is that the parameters then become physically meaningful and could be used, for example, to predict the Hansen solubility parameters.

4.2 The group contribution concept

The group contribution concept is based on the idea that all molecules can be broken down into functional groups and these functional groups affect certain properties of a molecule in an additive manner independent of the other groups present in the molecule. B' for example is represented in this work as follows:

$$B' = A + \sum_{i=1}^m v_i dB_i \quad (4-5)$$

where v_i is the frequency (or number of occurrences) of group i and dB_i is its contribution. As shown in the sections following not all groups conform to this general scheme and suitable modifications had to be made. For D' the optimum contribution scheme was found to depend on the number of heavy atoms n_i :

$$D' = D + \sum_{i=1}^m v_i \frac{dE_i}{n_i} \quad (4-6)$$

4.3 The group interaction concept

The group contribution concept can sometimes be inadequate to describe the properties of multi functional compounds as the assumption of group additivity does not always hold. For this reason the idea of group interaction was developed (Nannoolal et al.⁶). For non-additive groups (typically hydrogen bonding groups) the value of B' is calculated in the following way (where Gl_{i-j} is the interaction between group i and j):

$$B' = A + \sum_{i=1}^m v_i dB_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n Gl_{i-j} \quad (4-7)$$

whereby the interaction of a group with itself is set to zero (the group contribution accounts for this) and $Gl_{i-j}=Gl_{j-i}$. Consider the following example of a compound with two OH (the numbers is superscript are to differentiate between them) and one NH_2 group. The double summation term in Eq. (4-7) results in $2*Gl_{OH-NH_2} + 1*Gl_{OH-OH}$.

	OH ⁽¹⁾	OH ⁽²⁾	NH ₂
OH ⁽¹⁾	0	1	1
OH ⁽²⁾	1	0	1
NH ₂	1	1	0

4.4 New group contribution approach

The correct definition of structural groups plays an important role in the development of a successful group contribution method. The groups should be simple to allow broad applicability but at the same time capture all significant effects on the property to be predicted. Along with the different groups, structural correction groups differentiate between isomers or capture further effects that are not limited to individual groups. Great care must be taken to ensure that there actually is a need to include a correction, as excess groups, while slightly improving the property correlation for the training set, in most cases lead to poor or even very erroneous prediction results outside this set.

In order to reduce the number of structural groups, unnecessarily bulky groups were in several cases split into separate groups. This means that only half the number of different groups is required to represent the larger groups of the Nannoolal method. In addition, it also allows for more compounds to be fragmented and reduces the need for more specific groups. In the case of a double bonded carbon (alkene) the previous approach employed six different groups to represent all the possible combinations while the new approach only needs three (see Figure 4.8). An additional advantage is that the new groups are now each backed up by more experimental data as they are present in a larger variety of molecules.

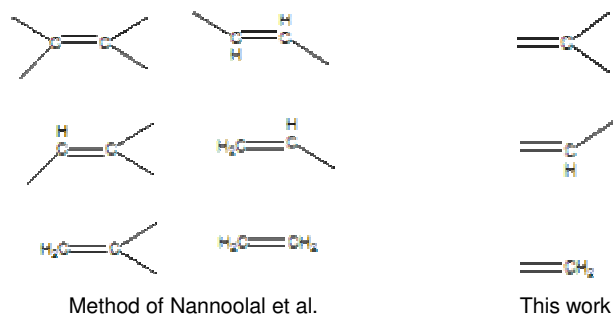


Figure 4.8 Comparison of the group contribution approaches for the non-cyclic alkene groups

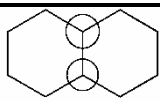
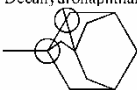
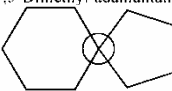
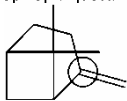
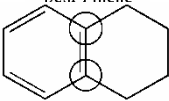
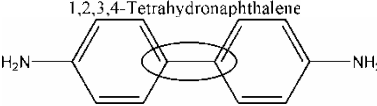
5 RESULTS AND DISCUSSION

5.1 Hydrocarbon compounds

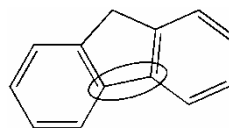
As mentioned earlier the new C-parameter employed in this work leads to a better representation of data for molecules of different sizes, as illustrated in Figure 5.1 to Figure 5.4. These figures show the performance of the vapour pressure equation (B' was regressed to the experimental data) using the C-parameter correlation of Nannoolal et al. compared to the new C-correlation (Eq.(4-3)). While both models perform adequately in the case of octadecane, only the new correlation is at the same time sufficiently suitable for propane.

Analysis of the hydrocarbon vapour pressure regressions revealed that data for some of the more complex compounds were not very well represented by the groups that were already in use and therefore the more specific structural groups, shown in Table 5.1, had to be added to account for several structural effects.

Table 5.1 New hydrocarbon structural groups (Ink No – fragmentation group number, Ref No – reference number is used to arrange like groups (e.g. halogen groups etc) since the ink no's have no real structure)

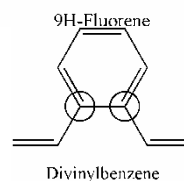
Ink No	Description	Example	Ref No
131	CH connecting two rings bonded to a carbon also connecting these rings	 cis-Decahydronaphthalene	113
136	Ring carbon attached to 3 other ring carbons and a sidechain carbon	 1,3-Dimethyl adamantane	114
137	Ring carbon attached to 4 other ring carbons	 Spiro[4.5]decane	115
134	Carbon in a ring double bonded to a sidechain carbon	 beta-Pinene	128
132	Aromatic carbon bonded to an aromatic carbon in a ring	 1,2,3,4-Tetrahydronaphthalene	205
133	2 aromatic carbon connected outside the rings	 Benzidine	206

135 Aromatic carbon connected to an aromatic carbon in a ring



207

138 Aromatic carbon connected to a double bonded carbon



208

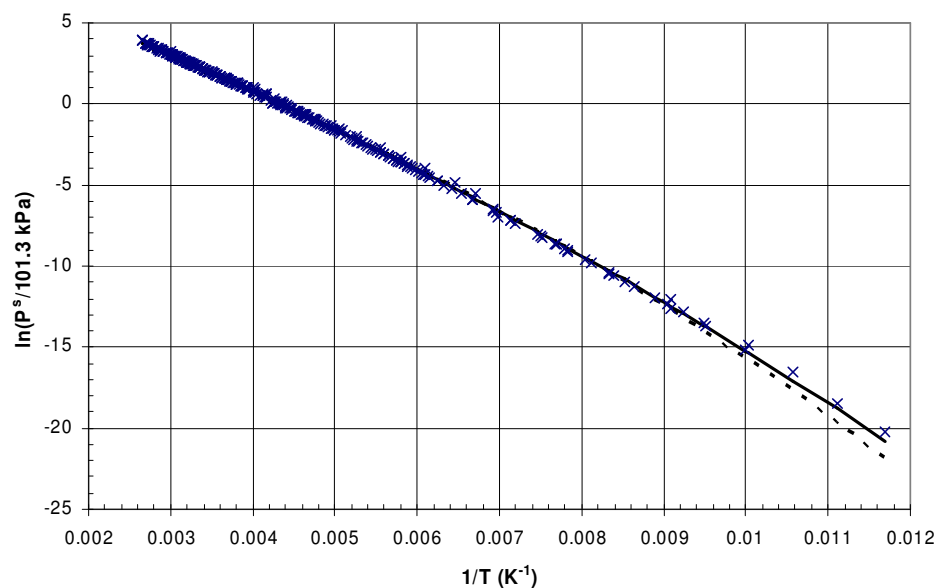


Figure 5.1 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for propane (x – data taken from the DDB², - - - data regressed using the C-parameter correlation of Nannoolal et al.⁸, — data regressed using the improved C-parameter correlation)

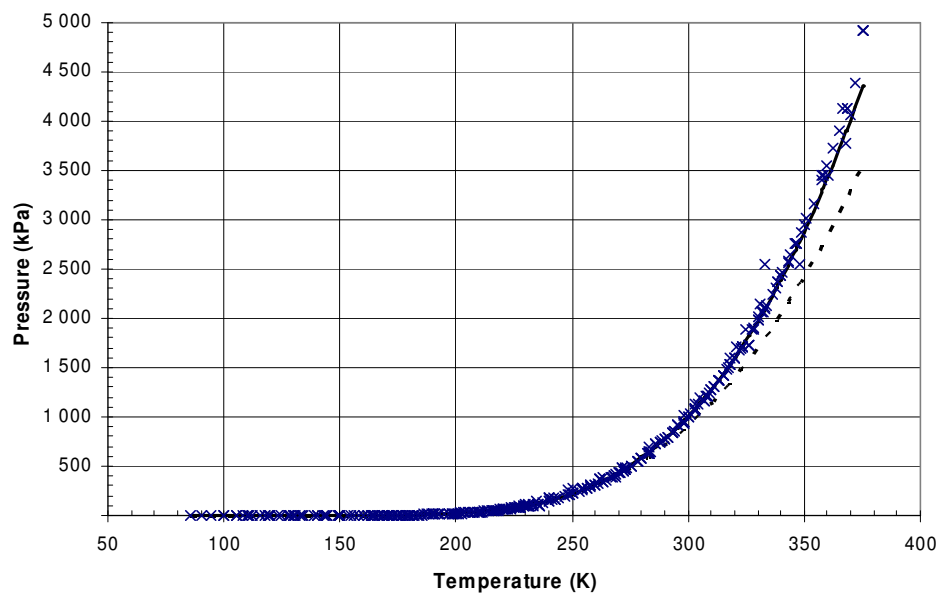


Figure 5.2 P^s vs. T for propane (x – data taken from the DDB², - - - data regressed using the C-parameter correlation of Nannoolal et al.⁸, — data regressed using the improved C-parameter correlation)

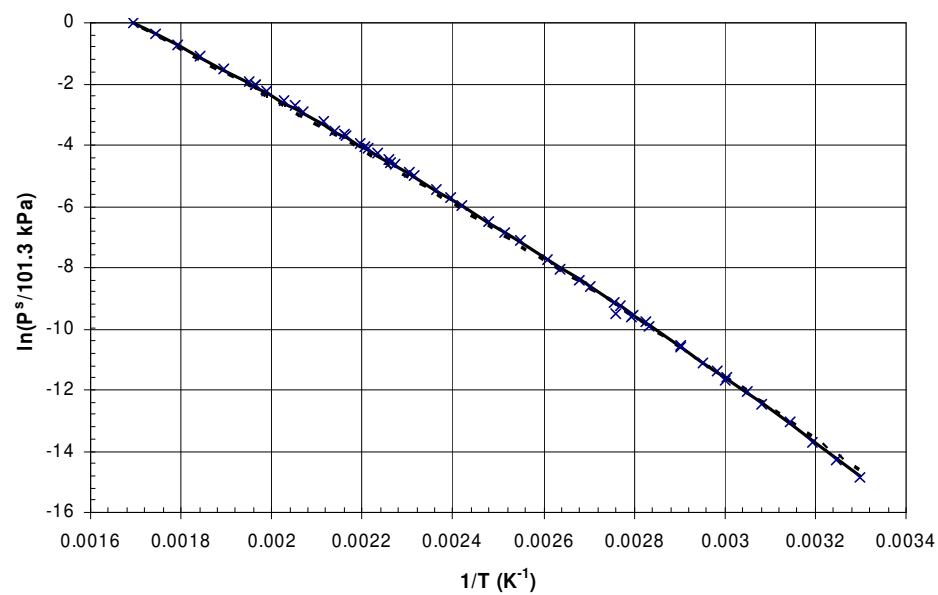


Figure 5.3 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for octadecane (x – data taken from the DDB², - - - data regressed using the C-parameter correlation of Nannoolal et al.⁸, — data regressed using the improved C-parameter correlation)

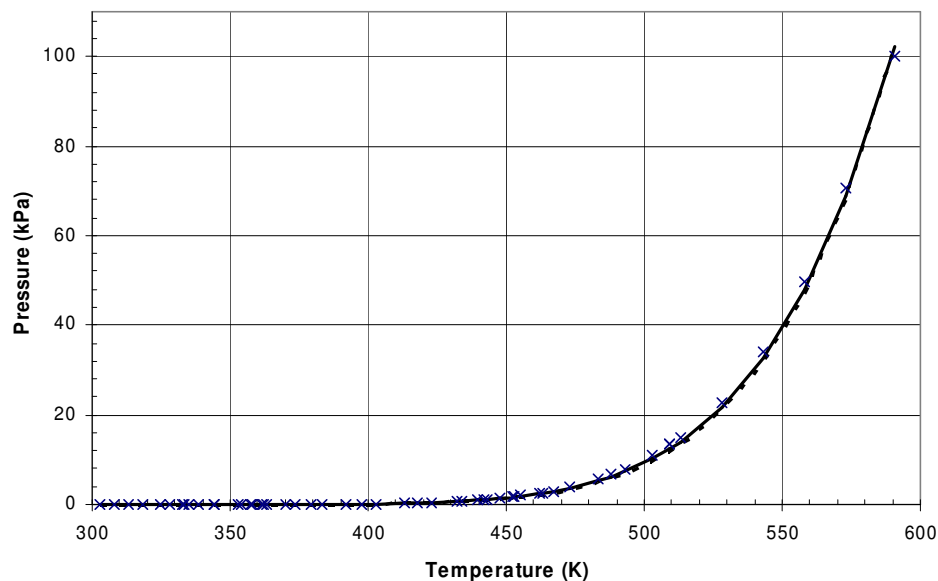
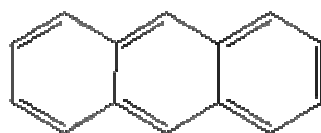


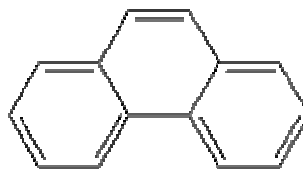
Figure 5.4 P^s vs. T for octadecane (x – data taken from the DDB², - - - data regressed using the C-parameter correlation of Nannoolal et al.⁸, — data regressed using the improved C-parameter correlation)

One of the problems sometimes encountered by group contribution methods is the inability to distinguish between isomers. With the hydrocarbons the difference between indistinguishable isomers (insofar as this method is concerned) is not very great. This difference is not very noticeable in either the boiling points or the B' parameter that was fitted, as an example consider the case of anthracene and phenanthrene:



Anthracene

$T_b = 613.2 \text{ K}$
 $B' = 9.538$



Phenanthrene

$T_b = 610.7 \text{ K}$
 $B' = 9.454$

Figure 5.5 Comparison of the properties of anthracene and phenanthrene.

Hydrocarbons provide the “backbone” for all of the other organic compounds. Special attention was paid to representing the behaviour of hydrocarbon compounds because any shortcomings would negatively affect the results for many other components. It became apparent during parameter regression that both small and large alkene and alkyne molecules exhibited larger than expected

deviation from group contribution prediction. Further analysis revealed that the B' values for these components were dependant on the number of atoms present in the molecule (Figure 5.6 & Figure 5.7).

The dB_i values were calculated by taking the difference between the fitted B' values for the n-alkenes (resp. n-alkynes) and the n-alkanes. The reason for this is that there is a large amount of reliable data for these species; also it is important to compare like with like in order to draw any meaningful conclusions from the data. In order to account for the size dependence a new group (with a frequency of 1) was added to all alkene and alkyne compounds and the following size dependant contribution scheme was used:

$$dB = \sum_i v_i dB_i + n_a \sum_j v_j dB_j + \sum_k dB_k \quad (5-1)$$

The subscript i covers all normal (size independent) groups, the subscript j covers all size dependant (e.g. alkene) groups and subscript k is for the size dependant group constants and therefore does not have a frequency term. For example if a molecule has 3 alkene and 2 alkyne groups there will be only be two size dependant groups, one for the alkene groups and one for the alkyne groups (i.e. $\sum_k dB_k = dB_{alkyne} + dB_{alkene}$).

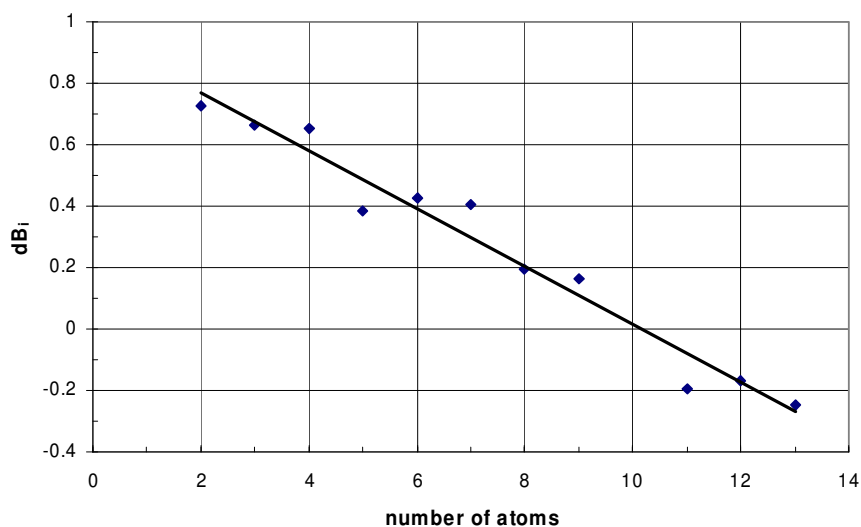


Figure 5.6 dB_i vs. number of atoms for different alkynes (♦ – dB_i data for each compound, — a linear least squares fit)

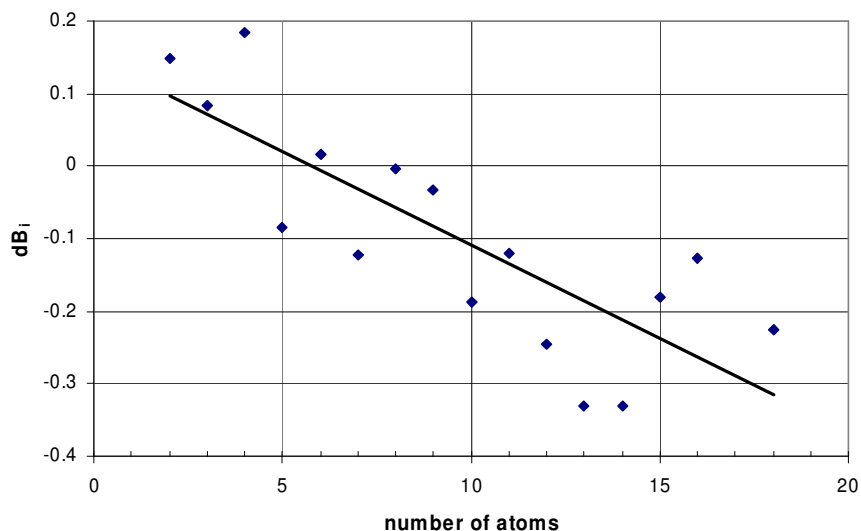


Figure 5.7 dB_i vs. number of atoms for different alkenes (\blacklozenge – dB_i data for each compound, — a linear least squares fit)

An average percentage error of 4.1% for the vapour pressure was obtained for the hydrocarbons which is a significant improvement compared to 5.4% obtained by Nannoolal et al.⁸. The mean relative deviations for various types of hydrocarbons and both methods are given in Table 5.2 & Table 5.3 for the different pressure regions.

Table 5.2 Relative mean deviation [%] in vapour pressure estimation for different types of hydrocarbons (this work).

The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
Hydrocarbons	532	23.5 ¹¹⁰²	6.8 ⁷⁹¹²	1.7 ²⁰⁵⁹²	5.0 ⁴⁶⁷⁷	4.1 ³⁴²⁹⁴
Alkanes	149	25.1 ⁷³⁸	8.8 ³⁴⁷⁵	2.1 ⁹⁰²³	6.4 ²⁸⁸³	5.4 ¹⁶¹²³
Non-cyclic alkanes	101	25.8 ⁶⁹⁷	9.8 ²⁸⁸⁴	2.4 ⁶⁸⁷³	6.7 ²⁶³⁸	6.1 ¹³⁰⁹⁵
Cyclic alkanes	48	13.0 ⁴¹	4.0 ⁵⁹¹	1.1 ²¹⁵⁰	3.1 ²⁴⁵	2.0 ³⁰²⁸
Alkenes	158	19.3 ⁴⁵	5.2 ⁶⁵⁸	1.5 ³¹⁶⁸	3.3 ⁷²⁰	2.5 ⁴⁵⁹¹
Non-cyclic alkenes	120	19.3 ⁴⁵	5.7 ⁵¹⁴	1.5 ²⁵¹³	3.2 ⁷⁰⁰	2.6 ³⁷⁷²
Cyclic alkenes	30	-	3.7 ¹⁰⁸	1.6 ⁵⁶⁵	4.4 ²⁰	2.0 ⁶⁹³
Alkynes	34	-	3.7 ¹³³	1.7 ⁶⁶⁶	2.5 ⁷⁹	2.1 ⁸⁷⁸
Non-cyclic alkynes	34	-	3.7 ¹³³	1.7 ⁶⁶⁶	2.5 ⁷⁹	2.1 ⁸⁷⁸
Aromatic hydrocarbons	140	20.4 ³¹²	5.4 ³¹⁹⁷	1.5 ⁶²⁹¹	2.1 ⁹⁷⁷	3.3 ¹⁰⁷⁸⁴

Table 5.3 Relative mean deviation [%] in vapour pressure estimation for different types of hydrocarbons (Nannoolal et al.). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
Hydrocarbons	527	51.4 ¹¹⁰²	9.0 ⁷⁷⁴⁴	1.7 ²⁰²⁵⁵	4.2 ⁴⁴⁵⁵	5.4 ³³⁵⁶⁷
Alkanes	149	62.6 ⁷³⁸	12.1 ³⁴⁷⁵	1.8 ⁹⁰²³	4.7 ²⁸⁸³	7.3 ¹⁶¹²³
Non-cyclic alkanes	101	63.2 ⁶⁹⁷	13.5 ²⁸⁸⁴	2.0 ⁶⁸⁷³	5.0 ²⁶³⁸	8.4 ¹³⁰⁹⁵
Cyclic alkanes	48	51.4 ⁴¹	5.5 ⁵⁹¹	0.9 ²¹⁵⁰	2.3 ²⁴⁵	2.6 ³⁰²⁸
Alkenes	154	50.3 ⁴⁵	6.4 ⁴⁹⁰	1.5 ²⁸⁷²	3.0 ⁵⁴⁰	2.9 ³⁹⁴⁷
Non-cyclic alkenes	116	50.3 ⁴⁵	7.4 ³⁴⁶	1.5 ²²¹⁷	3.0 ⁵²⁰	3.1 ³¹²⁸
Cyclic alkenes	30	-	4.3 ¹⁰⁸	1.8 ⁵⁶⁵	2.2 ²⁰	2.2 ⁶⁹³
Alkynes	33	-	4.9 ¹³³	2.0 ⁶²⁵	9.6 ³⁷	2.8 ⁷⁹⁵
Non-cyclic alkynes	33	-	4.9 ¹³³	2.0 ⁶²⁵	9.6 ³⁷	2.8 ⁷⁹⁵
Aromatic hydrocarbons	140	25.8 ³¹²	6.5 ³¹⁹⁷	1.7 ⁶²⁹¹	3.0 ⁹⁷⁷	3.9 ¹⁰⁷⁸⁴

The greatest improvement was found for the aliphatic hydrocarbons, where the error at ELP (extremely low pressure < 10 Pa) is significantly improved from the method of Nannoolal et al.⁸. The errors at high pressures are slightly worse than the previous method, this is to be expected because as shown in paragraph 2.2.1.1 the Antoine equation is deficient at high pressures and therefore there must be a trade-off between the high and the low pressure errors. This slight decrease in the error is compensated for with the large improvement of the low pressure errors.

5.2 Oxygen compounds

Oxygen compounds exhibited the largest deviation in case of both the component specific and group specific regressions. The largest deviations were observed for aliphatic alcohols and carboxylic acids. For this reason a logarithmic correction term (Eq.(4-4)) was added to properly model the data. Figure 5.8 & Figure 5.9 show the difference between results with and without this modification in case of 1-nonanol. A similar kind of deviation was observed for carboxylic acids and is shown for palmitic acid in Figure 5.10 & Figure 5.11. As mentioned above (see paragraph 4.1) this logarithmic correction term not only provides a better representation of the vapour pressure curve but also makes the model more physically realistic.

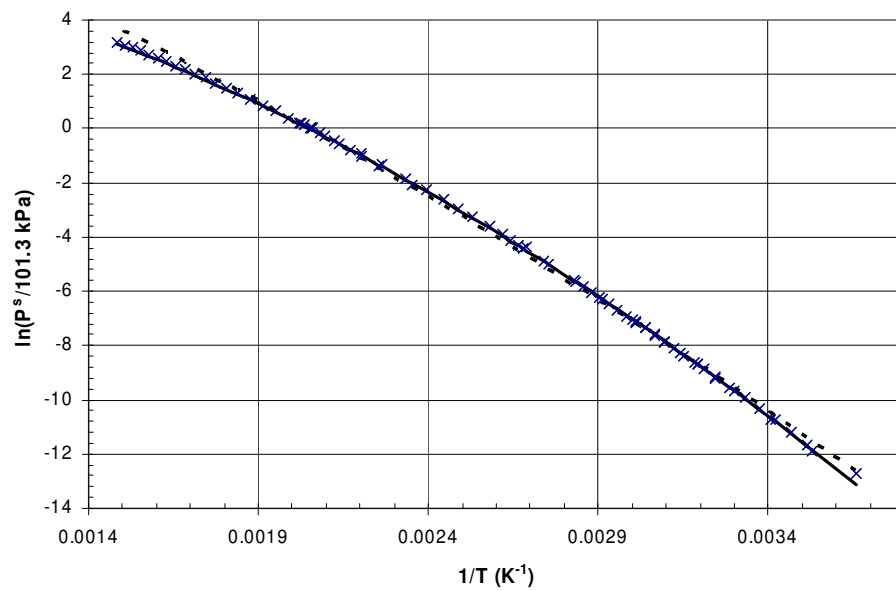


Figure 5.8 $\ln(P^s/101.3\text{ kPa})$ vs. $1/T$ for 1-nonanol (x – data taken from the DDB², - - - data regressed using the method of Nannoolal et al.⁸, — data regressed using the new logarithmic correction)

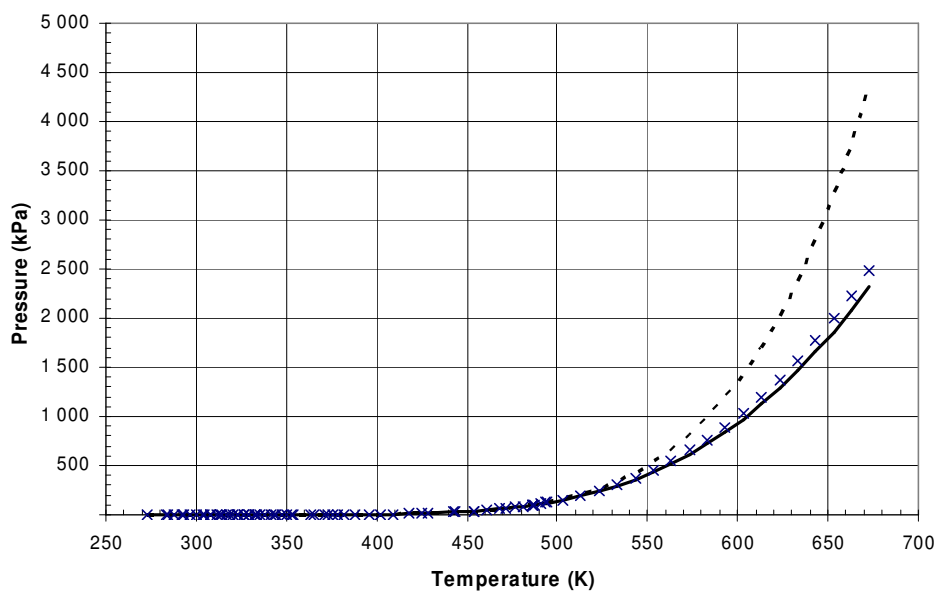


Figure 5.9 P^s vs. T for 1-nonanol (x – data taken from the DDB², - - - data regressed using the method of Nannoolal et al.⁸, — data regressed using the new logarithmic correction)

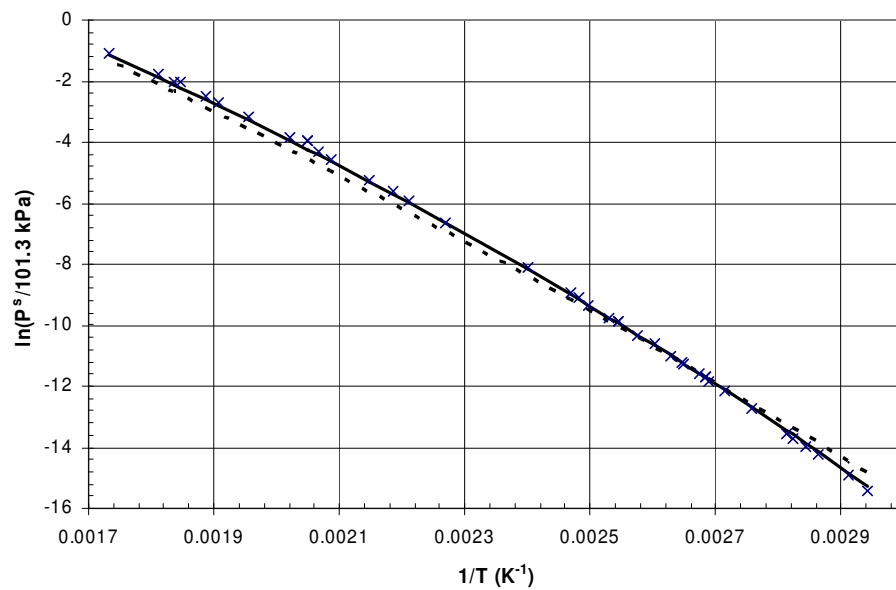


Figure 5.10 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for palmitic acid (x – data taken from the DDB², - - - data regressed using the method of Nannoolal et al.⁸, — data regressed using the new logarithmic correction)

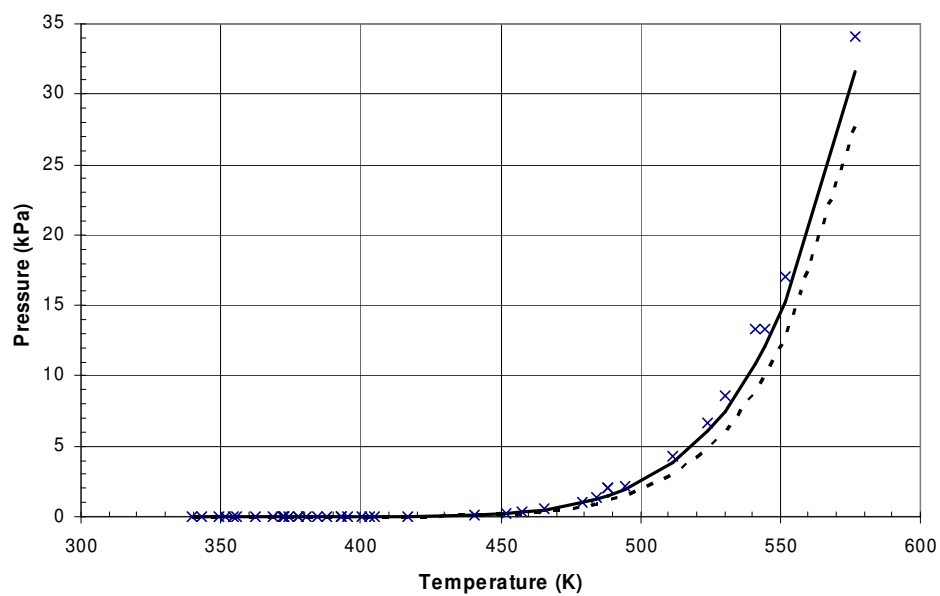


Figure 5.11 P^s vs. T for palmitic (x – data taken from the DDB², - - - data regressed using the method of Nannoolal et al.⁸, — data regressed using the new logarithmic correction)

As in case of the hydrocarbon group contributions, several types of oxygen containing components could not be represented by a size independent contribution (Eq. (4-5)) alone. Figure 5.12 &

Figure 5.13 show the effect of molecular size on dB_i (calculated in a similar way as in case of alkenes and alkynes before).

For aliphatic carboxylic acids and aliphatic alcohols there was a total change in the behaviour of dB_i when going from small to large molecules. This effect was accounted for by two separate groups for large and small molecules and both followed a similar scheme as for alkenes and alkynes (i.e. size dependant groups). The graphical representations of the dB_i values as function of molecular size are shown in Figure 5.14 & Figure 5.15. The quite noticeable deviation of the aliphatic alcohols and aliphatic carboxylic acids is more than likely due to hydrogen bonding, however it is unclear why the their aromatic counterparts do not exhibit the same effect.

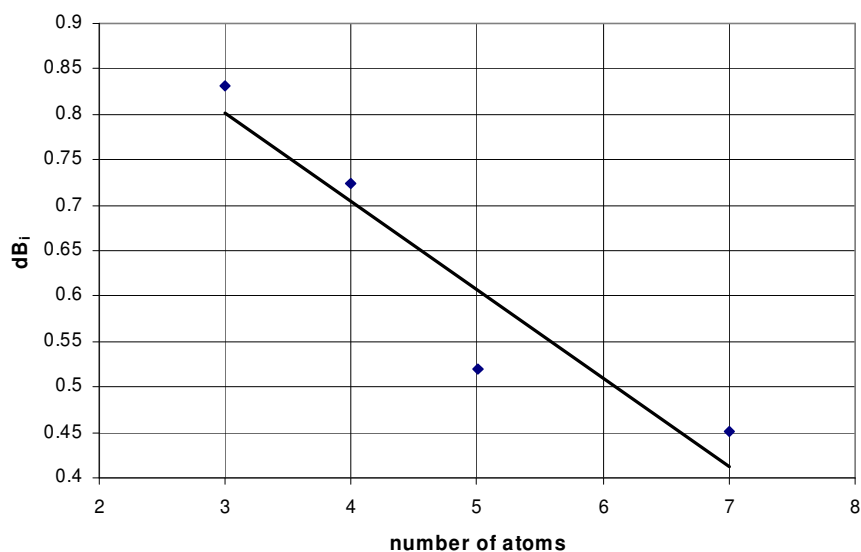


Figure 5.12 dB_i vs. number of atoms for different epoxides (♦ – dB_i data for each compound, — a linear least squares fit)

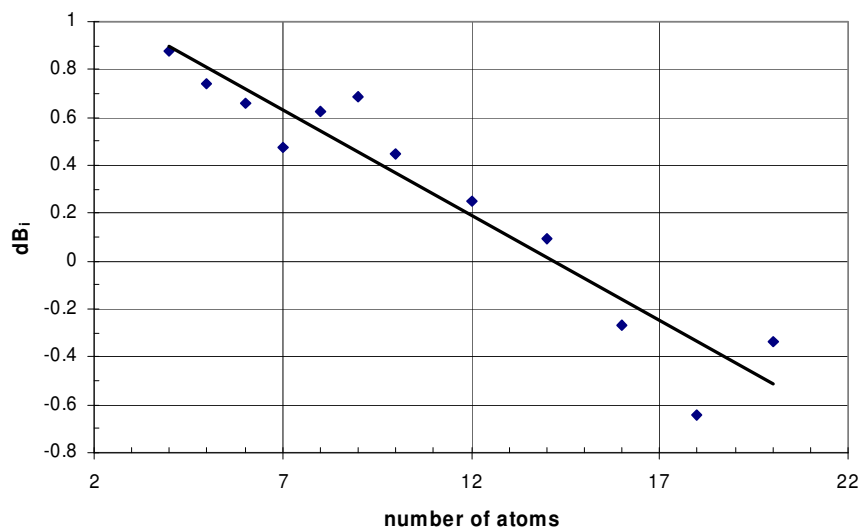


Figure 5.13 dB_i vs. number of atoms for different ketones (\blacklozenge – dB_i data for each compound, — a linear least squares fit)

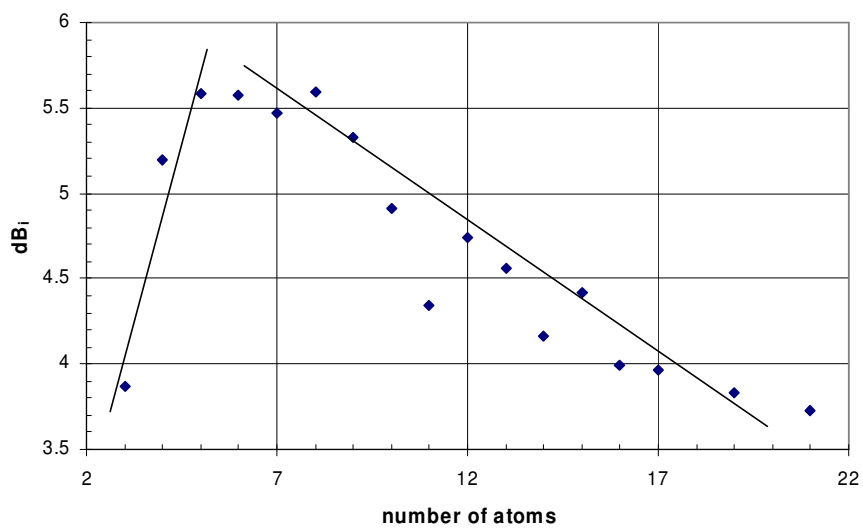


Figure 5.14 dB_i vs. number of atoms for different aliphatic alcohols (\blacklozenge – dB_i data for each compound, — lines to show the trends)

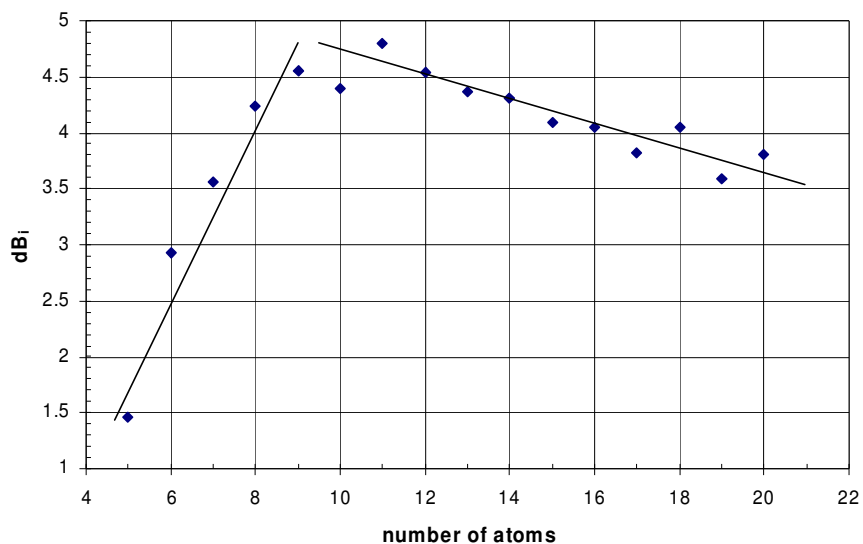


Figure 5.15 dB_i vs. number of atoms for different aliphatic carboxylic acids (♦ – dB_i data for each compound, — lines to show the trends)

The largest errors were observed for multifunctional aliphatic alcohols (diols, triols etc.). Closer analysis showed that the group interactions were dependant on the size of the molecule in a similar way as the contribution of the OH group. The OH-OH group interaction contributions calculated from the B' values of individual components were plotted against molecule size (Figure 5.16).

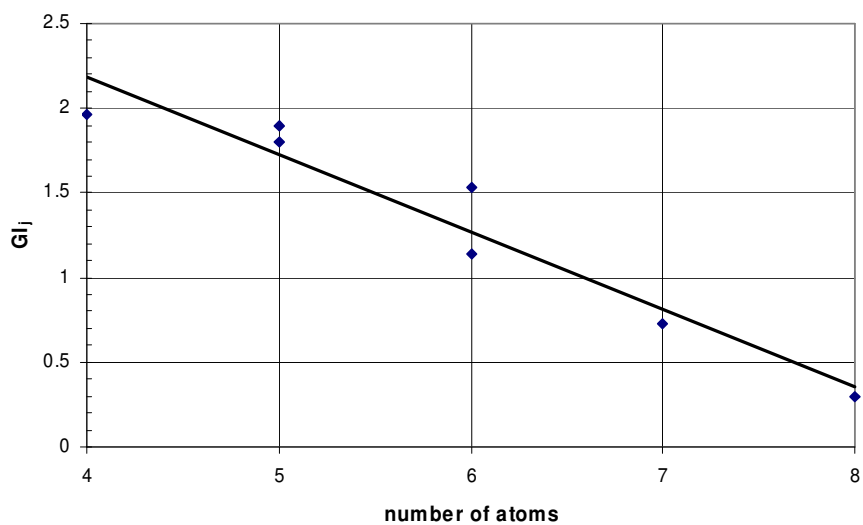


Figure 5.16 GI_i vs. number of atoms for different diols (♦ – dB_i data for each compound, — a linear least squares fit)

The relative mean deviation for all oxygen containing compounds was 6.5 % (36701 data points for 638 compounds). This compares well with the 8.5 % (36450 data points for 618 compounds) achieved by Nannoolal et al.⁸. There was a considerable improvement for the aliphatic alcohols, aliphatic carboxylic acids and ketones. As with the hydrocarbons the bulk of this improvement is found for the low pressure data. For some species (anhydrides and for some aromatic alcohols) there is actually a decrease in the performance. This is because while the new C-parameter is better for most compounds there are a few where it is a bit worse. This is however unavoidable and, considering the large improvement for the many other species, inconsequential. Overall the oxygen compounds show considerable improvement in the LP and ELP regions with reasonable improvement in the MP and HP regions.

Table 5.4 Relative mean deviation [%] in vapour pressure estimation for different types of oxygen containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All oxygen compounds	638	27.3 ⁸⁸²	12.4 ¹⁰⁹¹⁹	2.9 ²²⁷⁰⁸	6.3 ²¹⁷⁵	6.5 ³⁶⁷⁰¹
Carboxylic acids	35	23.9 ¹¹⁶	12.2 ¹⁸²³	3.8 ¹¹⁰⁶	13.8 ¹⁰	9.6 ³⁰⁵⁶
Aromatic carboxylic acids	1	-	1.6 ¹¹¹	2.7 ²²	-	1.8 ¹³³
Aliphatic carboxylic acids	34	23.9 ¹¹⁶	12.8 ¹⁷¹²	3.8 ¹⁰⁸⁴	13.8 ¹⁰	9.9 ²⁹²³
Alcohols	167	28.3 ³¹²	17.2 ³⁷⁶¹	4.1 ⁷¹⁸⁶	10.8 ⁶²⁶	9.2 ¹¹⁸⁸⁶
Aromatic alcohols	55	39.9 ²⁶	25.1 ⁶⁷⁸	3.1 ¹²⁷⁷	6.9 ⁶⁹	11.0 ²⁰⁵⁰
Aliphatic alcohols	112	27.3 ²⁸⁶	15.5 ³⁰⁸³	4.3 ⁵⁹⁰⁹	11.3 ⁵⁵⁷	8.8 ⁹⁸³⁶
Ethers	92	23.7 ¹⁷	6.5 ⁷⁷¹	1.8 ³⁷⁴³	3.4 ⁷³⁸	2.8 ⁵²⁶⁹
Esters	158	29.0 ³²⁶	8.4 ²⁴⁴⁵	2.2 ⁵⁶²²	5.6 ⁴⁰³	5.1 ⁸⁸⁰⁹
Ketones	64	19.0 ⁷⁵	8.9 ⁸⁸⁶	1.9 ²⁶³⁷	4.4 ²⁷¹	4.0 ³⁸⁶⁹
Aldehydes	30	6.1 ⁴	12.1 ³⁹³	2.4 ⁷³⁶	5.1 ¹⁹	5.7 ¹¹⁵³
Carbonate diesters	3	-	1.8 ³⁰	0.8 ²⁸⁹	-	0.9 ³¹⁹
Anhydrides	6	-	14.9 ⁵⁴	4.4 ⁸¹	-	8.6 ¹³⁵
Epoxides	11	-	2.9 ³³	1.8 ²⁸⁰	2.3 ⁴³	1.9 ³⁵⁶
Carbonates	3	38.7 ³	9.1 ¹⁵⁰	3.2 ⁹²	4.4 ¹⁰	7.2 ²⁵⁵
Ureas	5	14.8 ³	10.5 ⁵⁹	2.2 ⁴⁰	-	7.4 ¹⁰²

Table 5.5 Relative mean deviation [%] in vapour pressure estimation for different types of oxygen containing compounds (Nannoolal et al.⁸). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All oxygen compounds	618	60.1 ⁸⁸²	15.4 ¹⁰⁸¹⁰	3.2 ²²⁵⁴⁴	7.9 ²¹⁵¹	8.5 ³⁶⁴⁰⁴
Carboxylic acids	34	106.8 ¹¹⁶	17.2 ¹⁸¹⁹	3.5 ¹¹⁰⁰	10.6 ¹⁰	15.7 ³⁰⁴⁶
Aromatic carboxylic acids	1	-	13.7 ¹¹¹	2.4 ²²	-	11.8 ¹³³
Aliphatic carboxylic acids	33	106.8 ¹¹⁶	17.5 ¹⁷⁰⁸	3.6 ¹⁰⁷⁸	10.6 ¹⁰	15.8 ²⁹¹³
Alcohols	162	80.8 ³¹²	21.7 ³⁷³²	4.7 ⁷¹¹⁹	13.0 ⁶⁰²	12.5 ¹¹⁷⁶⁶
Aromatic alcohols	55	33.5 ²⁶	24.5 ⁶⁷⁸	3.3 ¹²⁷⁷	6.3 ⁶⁹	10.8 ²⁰⁵⁰
Aliphatic alcohols	107	85.1 ²⁸⁶	21.0 ³⁰⁵⁴	5.0 ⁵⁸⁴²	13.9 ⁵³³	12.9 ⁹⁷¹⁶
Ethers	91	28.3 ¹⁷	7.6 ⁷⁶⁸	2.1 ³⁷³⁸	4.7 ⁷³⁸	3.3 ⁵²⁶¹
Esters	154	30.8 ³²⁶	9.0 ²⁴³²	2.3 ⁵⁶⁰²	7.0 ⁴⁰³	5.5 ⁸⁷⁷⁶
Ketones	63	47.2 ⁷⁵	11.8 ⁸⁸⁶	2.0 ²⁶³⁷	5.6 ²⁷¹	5.3 ³⁸⁶⁹
Aldehydes	29	20.9 ⁴	15.0 ³⁹²	2.3 ⁷³¹	5.1 ¹⁹	6.8 ¹¹⁴⁷
Carbonate diesters	3	-	4.5 ³⁰	0.8 ²⁸⁹	-	1.1 ³¹⁹
Anhydrides	6	-	8.5 ⁵⁴	3.9 ⁸¹	-	5.7 ¹³⁵
Epoxides	11	-	4.6 ³³	2.2 ²⁸⁰	2.4 ⁴³	2.4 ³⁵⁶
Carbonates	3	27.8 ³	10.8 ¹⁵⁰	6.3 ⁹²	14.0 ¹⁰	9.5 ²⁵⁵
Ureas	4	104.2 ³	14.9 ³⁹	1.4 ³⁸	-	11.8 ⁸⁰

5.3 Nitrogen compounds

Most nitrogen compounds did not show significant deviation from the model predictions. Only in the case of primary aliphatic amines, nitriles and aliphatic isocyanates similar extensions as in the case of some of the oxygenated compounds were required. Figure 5.17 to Figure 5.19 show the ΔB_i contributions for these compound classes as function of molecular size.

Differentiation had to be made between aliphatic and aromatic isocyanates. Groups were also added for cyclic tertiary amines and the hydrazine group (in accordance with the new group contribution approach mentioned in paragraph 4.4)

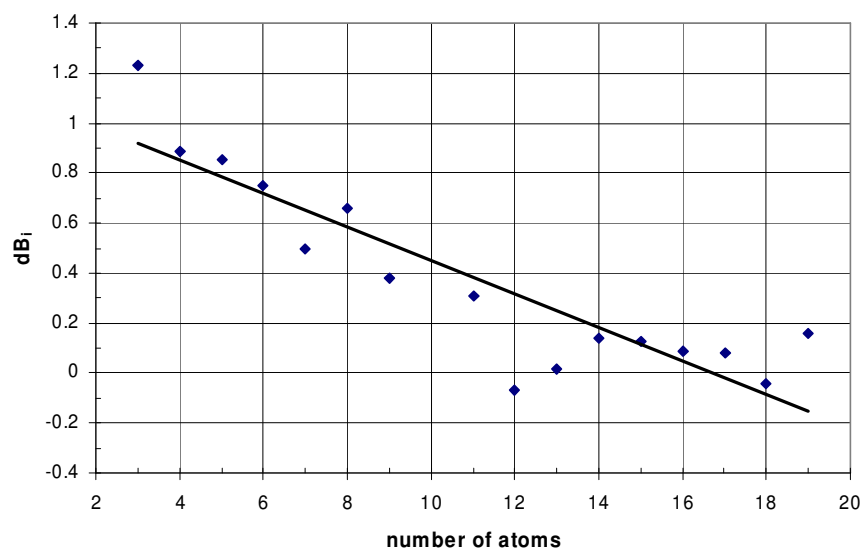


Figure 5.17 dB_i vs. number of atoms for different primary aliphatic amines (♦ – dB_i data for each compound, — a linear least squares fit)

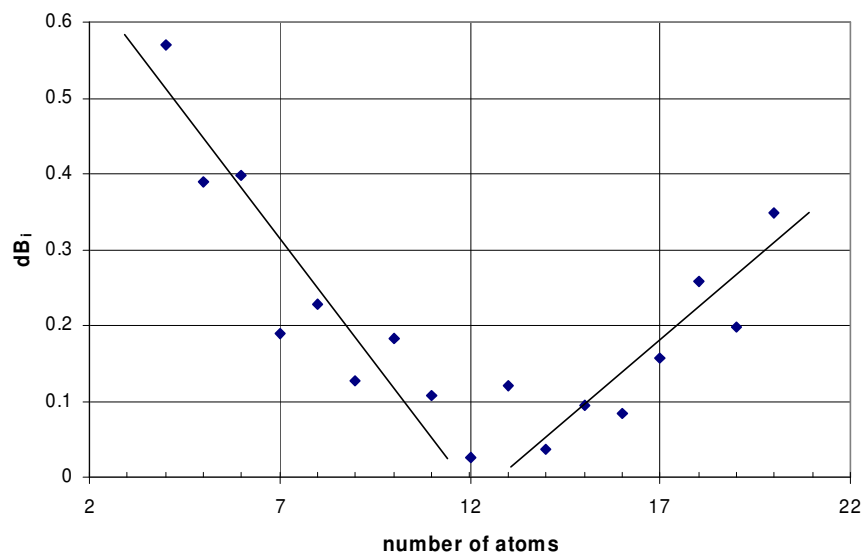


Figure 5.18 dB_i vs. number of atoms for different nitriles (♦ – dB_i data for each compound, — lines to show the trends)

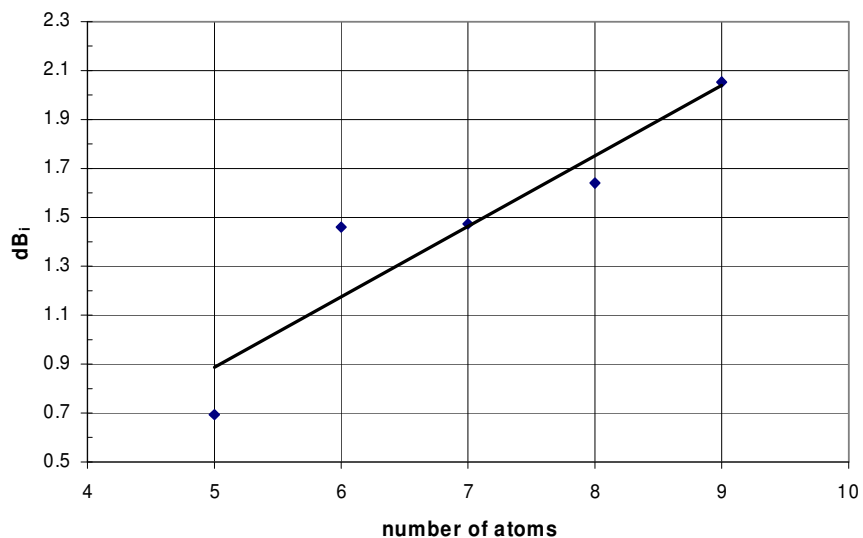


Figure 5.19 dB_i vs. number of atoms for different aliphatic isocyanates (♦ – dB_i data for each compound, — a linear least squares fit)

The relative mean deviation for all nitrogen containing compounds was 5.3% (10410 data points for 260 compounds) compared to 6.5% (10318 data points for 252 compounds) for the method of Nannoolal et al.⁸. It is fairly interesting to note that even with the special attention that was paid to the nitrile compounds there is actually a decrease in the performance of the model. As before the reason for this is that the C-parameter causes the nitriles to deviate quite significantly from group contribution. Even the two sets of size dependant groups could not properly account for this effect and therefore nitrile predictions should be used with a fair bit of caution when using temperatures far removed from the boiling point. The nitrogen compounds show a good improvement for all four pressure ranges with some compounds exhibiting considerable improvement (e.g. the nitrates)

Table 5.6 Relative mean deviation [%] in vapour pressure estimation for different types of nitrogen containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All nitrogen compounds	260	28.4 ⁸⁰	9.8 ³⁵²¹	2.5 ⁶³¹¹	5.0 ⁴⁹²	5.3 ¹⁰⁴¹⁰
Amides	18	12.8 ¹	11.6 ⁶³⁵	3.5 ⁵⁴⁹	8.3 ²	7.8 ¹¹⁹⁰
Isocyanates	6	4.5 ²	12.3 ⁵³	4.6 ⁶³	5.8 ⁵	7.9 ¹²³
Oximes	7	9.6 ⁵	5.5 ⁵⁷	1.9 ⁴⁹	-	4.1 ¹¹¹
Nitro groups	18	31.6 ²¹	7.7 ³³⁵	3.0 ⁴¹²	2.7 ¹²	5.8 ⁷⁸⁰
Nitrites	2	-	-	0.6 ³⁰	-	0.6 ³⁰
Azoles	1	-	-	0.6 ¹⁸	-	0.6 ¹⁸
Nitrates	6	79.9 ³	16.4 ⁴⁴	3.3 ²³	-	14.8 ⁷⁰
Amines	113	19.9 ⁹	12.1 ¹¹⁶⁹	2.6 ²⁶¹¹	6.3 ²⁹⁸	5.6 ⁴⁰⁸⁸
Primary amines	48	21.7 ³	13.7 ⁵⁶⁶	2.9 ¹¹⁸⁸	5.9 ¹²¹	6.4 ¹⁸⁷⁸
Secondary amines	33	17.9 ¹	12.8 ²⁸⁶	2.8 ⁸⁴²	6.3 ¹²¹	5.4 ¹²⁵¹
Tertiary amines	25	19.2 ⁵	8.3 ²⁹⁵	1.8 ⁴⁷⁰	6.0 ²⁹	4.5 ⁷⁹⁹
Azenes	3	-	9.7 ²²	1.2 ²⁶	-	5.1 ⁴⁸
Aromatic nitrogens	41	18.7 ¹³	7.2 ⁶⁵⁷	1.8 ¹⁵³⁹	3.2 ⁸²	3.5 ²²⁹¹
Nitriles	35	31.5 ²⁸	6.9 ⁴⁸³	2.3 ⁸⁹³	2.9 ⁹⁸	4.4 ¹⁵⁰³
Hydrazines	7	-	6.0 ⁶³	2.9 ⁷²	-	4.3 ¹³⁵

Table 5.7 Relative mean deviation [%] in vapour pressure estimation for different types of nitrogen containing compounds (Nannoolal et al.⁶). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All nitrogen compounds	252	61.8 ⁸⁰	11.6 ³⁴⁸⁴	2.8 ⁶²⁵⁶	7.7 ⁴⁹²	6.5 ¹⁰³¹⁸
Amides	16	42.3 ¹	11.7 ⁶³¹	3.4 ⁵³³	3.9 ²	7.9 ¹¹⁷⁰
Isocyanates	5	92.9 ²	22.7 ⁵⁰	6.0 ⁶⁰	13.1 ⁵	14.9 ¹¹⁷
Oximes	7	8.1 ⁵	4.3 ⁵⁷	2.5 ⁴⁹	-	3.7 ¹¹¹
Nitro groups	17	162.8 ²¹	10.3 ³³²	2.7 ⁴⁰⁹	2.1 ¹²	10.3 ⁷⁷⁴
Nitrites	2	-	-	6.3 ³⁰	-	6.3 ³⁰
Azoles	1	-	-	0.6 ¹⁸	-	0.6 ¹⁸
Nitrates	6	152.9 ³	31.4 ⁴⁴	5.9 ²³	-	28.2 ⁷⁰
Amines	112	41.3 ⁹	14.5 ¹¹⁶⁸	3.0 ²⁶⁰⁶	9.0 ²⁹⁸	6.8 ⁴⁰⁸²
Primary amines	48	28.7 ³	13.4 ⁵⁶⁶	3.3 ¹¹⁸⁸	10.8 ¹²¹	6.9 ¹⁸⁷⁸
Secondary amines	33	83.5 ¹	17.3 ²⁸⁶	2.5 ⁸⁴²	5.5 ¹²¹	6.2 ¹²⁵¹
Tertiary amines	25	40.3 ⁵	13.8 ²⁹⁵	3.1 ⁴⁷⁰	5.8 ²⁹	7.4 ⁷⁹⁹
Azenes	0	-	-	-	-	-
Aromatic nitrogens	41	5.2 ¹³	8.2 ⁶⁵⁷	2.0 ¹⁵³⁹	4.0 ⁸²	3.8 ²²⁹¹
Nitriles	35	19.5 ²⁸	5.2 ⁴⁸³	1.9 ⁸⁹³	7.4 ⁹⁸	3.7 ¹⁵⁰³
Hydrazines	7	-	45.4 ⁶³	11.5 ⁷²	-	27.3 ¹³⁵

5.4 Sulfur compounds

Sulfur containing compounds showed no significant deviation in both single compound and group contribution regressions. The average error for all sulfur containing compounds was 3.5% (3386 data points for 104 compounds) compared to 11.1% (3378 data points for 103 compounds) obtained with the method of Nannoolal et al.⁸. This huge improvement in the overall error is misleading since the massive error for the sulfoxides significantly offsets this error. When the sulfoxides are removed from the error calculation, the error is slightly over 4% which is comparable to the current method. This shows that the sulfur compounds follow the principle of group contribution very well (a very good example of this is the mercaptans – sometimes known as thiols – where the overall percentage error for 37 different compounds is only 2.2% for both methods)

Table 5.8 Relative mean deviation [%] in vapour pressure estimation for different types of sulfur containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All sulfur compounds	104	22.6 ³²	7.5 ⁷⁶⁹	1.7 ²⁴²⁵	7.5 ¹⁶⁰	3.5 ³³⁸⁶
Disulfides	8	-	1.1 ⁷⁴	1.3 ²⁰⁵	2.2 ⁸	1.2 ²⁸⁷
Mercaptans	38	6.6 ¹	4.5 ¹⁰⁸	1.3 ⁷⁹⁷	8.9 ⁶⁸	2.2 ⁹⁷⁴
Thioether	50	23.9 ²⁵	9.9 ³¹⁷	1.6 ¹²⁷³	5.6 ⁷⁷	3.7 ¹⁶⁹²
Sulfones	1	19.9 ⁶	17.5 ⁵⁹	4.9 ²⁶	21.7 ⁷	14.6 ⁹⁸
Sulfon amides	3	-	11.4 ¹⁸	0.1 ¹	-	10.8 ¹⁹
Sulfoxides	1	-	4.2 ¹⁴¹	4.9 ¹²⁰	-	4.5 ²⁶¹
Isothiocyanates	3	-	4.4 ⁵²	5.7 ³	-	4.5 ⁵⁵

Table 5.9 Relative mean deviation [%] in vapour pressure estimation for different types of sulfur containing compounds (Nannoolal et al.⁸). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All sulfur compounds	103	29.2 ³²	37.9 ⁷⁶⁹	2.6 ²⁴²¹	7.2 ¹⁵⁶	11.1 ³³⁷⁸
Disulfides	8	-	2.5 ⁷⁴	1.4 ²⁰⁵	2.6 ⁸	1.7 ²⁸⁷
Mercaptans	37	8.3 ¹	3.4 ¹⁰⁸	1.4 ⁷⁹³	9.8 ⁶⁴	2.2 ⁹⁶⁶
Thioether	50	28.7 ²⁵	11.9 ³¹⁷	1.6 ¹²⁷³	5.5 ⁷⁷	4.1 ¹⁶⁹²
Sulfones	1	34.6 ⁶	16.9 ⁵⁹	2.8 ²⁶	7.1 ⁷	13.5 ⁹⁸
Sulfon amides	3	-	9.4 ¹⁸	0.1 ¹	-	8.9 ¹⁹
Sulfoxides	1	-	166.3 ¹⁴¹	23.1 ¹²⁰	-	100.4 ²⁶¹
Isothiocyanates	3	-	4.5 ⁵²	5.2 ³	-	4.6 ⁵⁵

5.5 Halogen compounds

Halogen compounds are unique in that there are many multifunctional compounds which in this case were not treated like the other multifunctional species. Instead of a group interaction term there are groups which account for compounds with 1, 2 or 3 halogen atoms attached to the same carbon and in that way do provide some sort of group interaction. No specific problems were observed with modelling and predicting the vapour pressure curves of these components.

For fluoro, chloro and bromo compounds sufficient data were available to regress all the group contributions. Only in the case of iodo compounds was a single group was used due to the lack of data. Therefore results for compounds containing iodine should be used with caution. The relative mean deviation for all halogen containing compounds is 3.3% (19465 data points for 317 compounds) which is similar to the 4.1% obtained by the method of Nannoolal et al.⁸ (19344 data points for 300 compounds). The greatest improvement is found for ELP's where the error for the current work is half the error obtained by the method of Nannoolal et al.⁸. This is due to the addition of a couple more halogen groups (see groups 42 and 117 in Table A.1 for examples) and the new C-parameter.

Table 5.10 Relative mean deviation [%] in vapour pressure estimation for different types of halogen containing compounds (this work). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All halogen compounds	317	37.5 ⁵⁹	7.0 ²³²⁹	1.8 ¹¹⁰⁴⁹	4.2 ⁶⁰²⁶	3.3 ¹⁹⁴⁶⁵
Fluorine compounds	86	-	8.7 ⁴⁸¹	1.9 ³⁸³⁶	5.0 ³⁸⁰⁵	3.8 ⁸¹²³
Chlorine compounds	108	16.6 ¹⁷	6.5 ⁹⁷³	1.6 ⁴¹⁷⁰	4.4 ⁶⁰⁸	2.8 ⁵⁷⁶⁸
Bromine compounds	42	47.1 ⁴⁰	6.2 ⁴⁴⁹	1.7 ⁷⁸¹	0.4 ¹⁵	4.7 ¹²⁸⁵
Iodine compounds	10	21.4 ²	11.4 ¹²⁸	2.0 ²⁵³	4.2 ⁵	5.3 ³⁸⁸

Table 5.11 Relative mean deviation [%] in vapour pressure estimation for different types of halogen containing compounds (Nannoolal et al.⁸). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All halogen compounds	300	76.9 ⁵⁹	8.2 ²²⁶⁶	2.0 ¹⁰⁹⁸¹	5.5 ⁶⁰²⁶	4.1 ¹⁹³³⁴
Fluorine compounds	84	-	8.3 ⁴⁸¹	2.2 ³⁸²⁴	6.4 ³⁸⁰⁵	4.5 ⁸¹¹¹
Chlorine compounds	98	39.6 ¹⁷	8.1 ⁹³⁹	2.1 ⁴¹³⁶	6.2 ⁶⁰⁸	3.6 ⁵⁷⁰⁰
Bromine compounds	42	95.7 ⁴⁰	9.2 ⁴⁴⁹	1.9 ⁷⁸¹	2.2 ¹⁵	7.4 ¹²⁸⁵
Iodine compounds	10	19.1 ²	9.4 ¹²⁸	1.9 ²⁵³	3.8 ⁵	4.5 ³⁸⁸

5.6 Other compounds

This category of species is very broad and has the highest potential for large errors due to the fact that only few data were available for each group. This is especially true for the organometallics where there was very limited data available and therefore predicted results should only be used as a rough guide. Some attention was paid to the silicon containing compounds and the silicon groups were expanded in a similar fashion to the carbon compounds. There are however much fewer silicon groups as there are no cyclic silicon chains and much less data were available to back up very differentiated groups. Four groups for halogen-substituted silicon were added. The relative mean deviations are given in the following tables:

Table 5.12 Relative mean deviation [%] in vapour pressure estimation for various other compounds. The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
Phosphorous compounds	9	10.1 ⁷	10.7 ⁴⁵	4.6 ⁴⁹	-	7.7 ¹⁰¹
Metals	18	-	6.0 ¹⁰⁹	2.7 ²⁷⁸	1.2 ¹⁴	3.6 ⁴⁰¹
Other compounds	13	-	7.5 ¹⁵⁰	4.4 ¹⁹⁰	-	5.8 ³⁴⁰
Silicon compounds	68	-	9.7 ³⁸³	2.5 ⁷⁴⁶	2.8 ¹³⁸	4.7 ¹²⁶⁷

Table 5.13 Relative mean deviation [%] in vapour pressure estimation for various other compounds (Nannoolal et al.⁸). The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
Phosphorous Compounds	8	15.8 ⁷	11.1 ⁴⁵	2.4 ⁴⁹	-	7.2 ¹⁰¹
Metals	18	-	7.6 ¹⁰⁹	2.2 ²⁷⁸	2.0 ¹⁴	3.7 ⁴⁰¹
Other Compounds	12	-	11.2 ¹⁴⁶	5.5 ¹⁸⁸	-	8.0 ³³⁴
Silicon Compounds	68	-	14.9 ³⁸³	2.7 ⁷⁴⁶	1.9 ¹³⁸	6.3 ¹²⁶⁷

5.7 Testing the method

In the preceding paragraphs (5.1-5.6) the percentage errors were shown for all data that was contained in the model training set. Therefore in order to be sure that the model is not simply well fitted to the training set an external test set of data was used to test the validity of the model. The test-set contained a wide range of data so as to provide a realistic measure of the model performance. The overall percentage error for the test set was 7.1 % for 2879 data points (160 compounds), this error is somewhat inflated because there was quite a large scatter present in the

test set data (as discussed in paragraph 3.2)

The test-set revealed that the model is somewhat deficient in predicting the vapour pressures of alcohol and carboxylic acid (both aliphatic) compounds that contain a large amount of halogen compounds. The reason for this is that the logarithmic correction term makes B' dependant on the number of atoms (since D' is dependant on the number of atoms and B' and D' are strongly intercorrelated); this change in B' cannot be properly predicted by group contribution. The reason for this is that adding a halogen atom to a molecule does not have the same effect as a carbon (or any other non-metal atom). Table 5.14 shows the percentage errors for the test set data, the percentage errors of the current method and the method of Nannoolal et al.⁸ are shown. All compounds seem to exhibit fairly similar percentage errors, with the sulfur compounds having the lowest overall error.

Table 5.14 Percentage errors for the test set data

Compound Class	Current Method		Nannoolal et al. ⁸	
	NC	Error%	NC	Error%
Hydrocarbons	59	6.5 ⁴⁸⁶	59	7.5 ⁴⁸⁶
Halogen compounds	14	7.0 ⁷³	14	7.1 ⁷³
Oxygen compounds	46	7.1 ¹⁹⁵²	44	8.5 ¹⁹³⁸
Nitrogen compounds	19	7.3 ¹⁷²	18	8.9 ¹⁶⁹
Sulfur compounds	3	5.6 ²⁴	3	11.0 ²⁴
All compounds	157	7.1 ²⁸⁷⁹	154	8.2 ²⁸⁵⁹

5.8 Solid vapour pressures

The average error for the solid vapour pressure points was 21.1 % (152 compounds 4080 data points). This percentage is deceptive in that a higher percentage for low pressure data is still a fairly low absolute deviation. This coupled with the fact that there is a large scatter among solid vapour pressure data and considering that many of the compounds were not part of the training set makes the 21% error acceptable. A good example of a "blind prediction" (compound not part of the training set) for the solid vapour pressure data is given in Figure 5.20 (even for this fairly good prediction there was a percentage error of 17 % due to the obvious scatter present in the data).

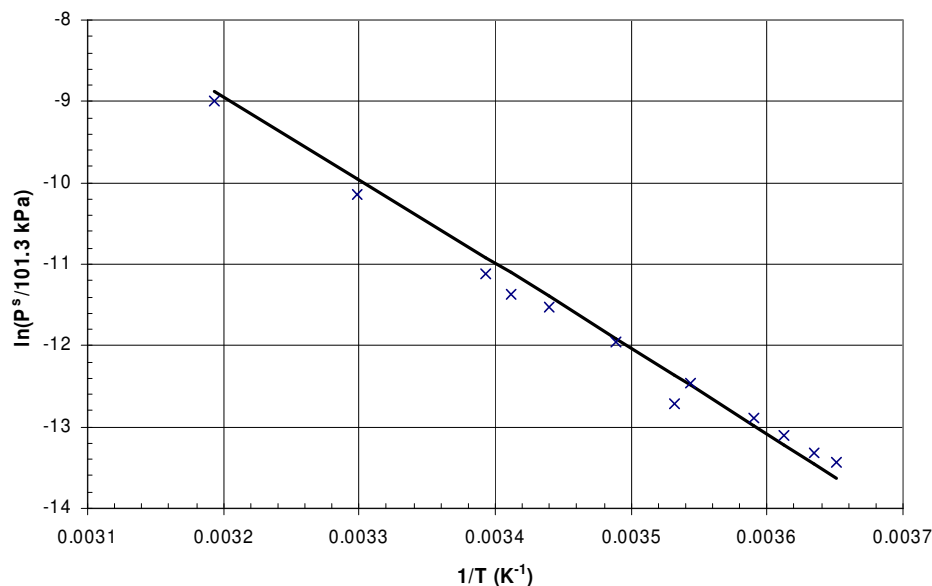


Figure 5.20 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for thymol (x – solid data taken from the DDB², — vapour pressure curve)

The disadvantage with using a method such as this to predict the solid vapour pressures is that the normal boiling point, melting point and heat of melting is required in order to make predictions. If there is no experimental normal boiling point it can be estimated (Rarey et al.⁵ and Nannoolal et al.⁶) and melting point data is quite widely available and simple to measure, however the heat of melting is rather more complex to measure and therefore a possible application could be the prediction of the heat of melting (or fusion) from experimental data. This would be a simple fit since the model is linear with respect to the heat of melting.

5.9 Heat of vaporization

The heat of vaporization is quite simply related to vapour pressure by the Clausius-Clapeyron equation (Eq. (2-12)). The heat of vaporization prediction for this model can therefore be given by the following expression:

$$\Delta H_{vap} = -R\Delta Z_{vap} \left[B' \frac{(C(T_b) - T_b)}{\left(1 - \frac{C(T_b)}{T}\right)^2} - D'T \right] \quad (5-2)$$

The problem with this equation is that it requires a value for ΔZ_{vap} which is difficult to compute

accurately (especially at high temperatures). However if only low temperatures are considered it can be assumed that ΔZ_{vap} is unity. This is a reasonable assumption since at lower temperatures (and therefore pressures) the gas and the liquid phases approximate ideal behaviour (i.e. $Z^v \rightarrow 1$ and $Z^l \rightarrow 0$). To test this assumption Eq. (5-2) was used to predict the heats of vaporization at 298.15 K. This test not only shows the validity of Eq. (5-2) but also serves as a good test of the physical realism of the model parameters (as discussed above).

Table 5.15 Percentage errors for heat of vaporization at 298.15 K using Eq. (5-2) with $\Delta Z = 1$

Compound Class	NC	Error%
Hydrocarbons	197	2.86
Halogen compounds	89	2.08
Oxygen compounds	224	4.19
Nitrogen compounds	101	2.92
Phosphorous compounds	1	6.93
Sulfur compounds	38	2.56
Metals	4	5.64
Other compounds	1	7.43
Silicon compounds	12	9.21
All compounds	718	3.48

The overall percentage error was found to be 3.48 %, Table 5.15 shows the percentage errors for each compound class. The prediction is very good for the hydrocarbon, halogen, nitrogen and sulfur compounds but is a bit poorer for the rest. The percentage error is very acceptable as the model was in no way fitted to heat of vaporization data and many of the compounds that were predicted were not in the training set for the vapour pressure model determination. This also shows that the model parameters must have some physical realism and therefore there should be no problem when predicting vapour pressures of compounds not contained in the training set.

5.10 Solubility parameters

A fairly novel (and indirect) application of vapour pressure data is the prediction of solubility. The relationship is not directly related but rather to heat of vaporization (paragraph 5.9). This relationship is given by calculating the cohesive energy unit per volume of a liquid as follows:

$$c = \frac{-U}{V} \approx \frac{\Delta U}{V} = \frac{\Delta H_{vap} - RT}{V_m} \quad (5-3)$$

This cohesive energy is an indication of how easy or difficult it is for molecules to escape the liquid. This is where the relationship between vaporization and solubility comes in, as in both cases molecules have to escape from one phase to another. Hildebrand (shown by Barton⁴⁸) suggested that the square root of the cohesive energy be used to describe solvency behaviour of compounds. The Hildebrand solubility parameter (δ) for this model can be given as ($V_m [=] \text{ cm}^3/\text{mol}$):

$$\delta = \left[-\frac{RT}{V_m} \left(\Delta Z_{vap} \left(B' \frac{T(C(T_b) - T_b)}{(T - C(T_b))^2} - D' \right) + 1 \right) \right]^{0.5} \quad (5-4)$$

As with the heat of vaporization, if a low enough temperature is used the change in compressibility factor can be assumed to be unity. Since there are no solubility parameters stored in the DDB² a sample set of compounds were used to show the accuracy of Eq. (5-4) (Table 5.16). For all the compounds used in the table there is good agreement between the predicted and the literature values of the Hildebrandt solubility parameters. Since the heat of vaporisation can be fairly well predicted for most species (see Table 5.15) it seems like there should be no real problem in predicting the Hildebrandt solubility parameters of these species.

Table 5.16 Results for the prediction of Hildebrand solubility parameters at 298 K

Name	Tb	δ^{298}	$\delta^{298} \text{ lit.}^{48}$
n-Hexane	342.1	15.1	14.9
Diethyl ether	307.6	15.7	15.4
1,1,1-Trichloroethane	347.3	17.3	17.5
Cyclohexane	353.9	16.7	16.8
Toluene	383.8	18.3	18.3
Benzene	353.3	18.6	18.7
Methyl ethyl ketone	352.8	19.1	19.3
Pyridine	388.6	21.4	21.7
n-Propanol	370.4	24.4	24.9

5.11 Advantage of group contribution

When experimental data are available for the component of interest, a regression of these data will always represent the experimental findings better than the group contribution estimation. However, the inherent advantage of group contribution is that it can help to identify unreliable data and in case of several differing data sets help to identify the more probable values. The reason for this is that the group contributions were regressed to a larger amount of data for a variety of components. This point is very well illustrated by the following example which was encountered during the development of the model. Figure 5.21 shows the vapour pressure plot for diethylmalonate.

Judging from the regression curve there is a significant scatter in the data. However the predicted curve reveals that the low pressure data must be erroneous

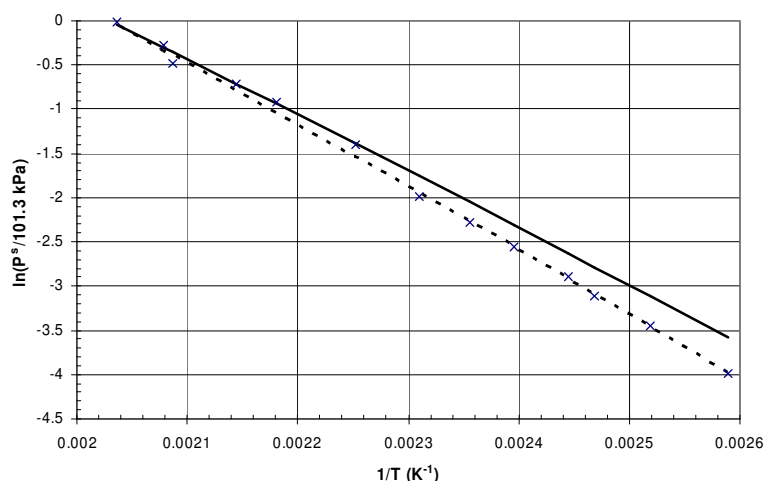


Figure 5.21 $\ln(P^s/101.3\text{kPa})$ vs. $1/T$ for diethyl malonate (x – data taken from the DDB², — predicted, - - - fitted)

5.12 General results and discussion

The vapour pressure percentage error for all compounds was found to be 5.0% (2332 compounds 113 888 data points) which compares very favourably to the method of Nannoolal which has an error of 6.6 % (2207 compounds 111 757 data points). As can be seen from Table 5.17 the greatest improvement is for pressures below 10 kPa (LP & ELP) where the percentage error is much lower than it was previously. For pressures above 10 kPa there is still an improvement but not nearly as noticeable as for the lower pressures. As the normal boiling temperature is supplied as a parameter, huge deviations in B' are required to produce larger errors in vapour pressure in the vicinity of atmospheric pressure. Therefore the improvement in case of low pressure data far away from the reference point is a sensible measure of model performance.

Table 5.17 Relative mean deviation [%] in vapour pressure estimation for the new method and the method of Nannoolal et al. The number in superscript is the number of data points used; the main number is the average percentage error of each data point. NC – Number of compounds; ELP – Extremely low pressure $P < 10$ Pa; LP – Low pressure $10 \text{ Pa} < P < 10 \text{ kPa}$; MP – Medium pressure $10 \text{ kPa} < P < 500 \text{ kPa}$; HP – High pressure $P > 500 \text{ kPa}$; AVE – Average error.

Group	NC	ELP	LP	MP	HP	AVE
All compounds (this work)	2332	25.7 ²²¹¹	9.8 ²⁸²³²	2.3 ⁶⁸⁷²⁹	5.0 ¹⁴⁶⁷⁸	5.0 ¹¹³⁸⁸⁷
All compounds (Nannoolal et al.)	2207	55.2 ²²⁰⁵	13.1 ²⁷⁶⁸⁷	2.6 ⁶⁷⁵¹⁸	5.7 ¹⁴³¹⁰	6.6 ¹¹¹⁷⁵⁷

In the sections preceding there has been very little mention of the chance of failure of the model (i.e. compounds with an unacceptably high vapour pressure – e.g. > 20% RMD). The reason for this is that there were only very few compounds which failed. Figure 5.22 shows that the vast majority of the data (85%) lies under the 10% RMD mark. The compounds with a RMD above 10% (or more specifically 10 – 20 %) are almost all there due to slightly lower quality data. Many compounds which are above the 20% RMD (4% of the total data) are a combination of suspect data (many have <10 data points from only one source) and genuine failure of the method. The best example of failure is with glycols (e.g. triethylene glycol). Even with the special attention that has been given to the alcohols, the glycols are still very difficult to predict accurately. Similarly, diols with other non – hydrocarbon groups (e.g. ethylene nitrate – 40% RMD) are also quite difficult to accurately predict (some attention was given to this in paragraph 5.7). While any error greater than 20% is not great it can still be useful if a rough vapour pressure estimate is needed, and it is useful to note that there were no errors exceeding 60% RMD.

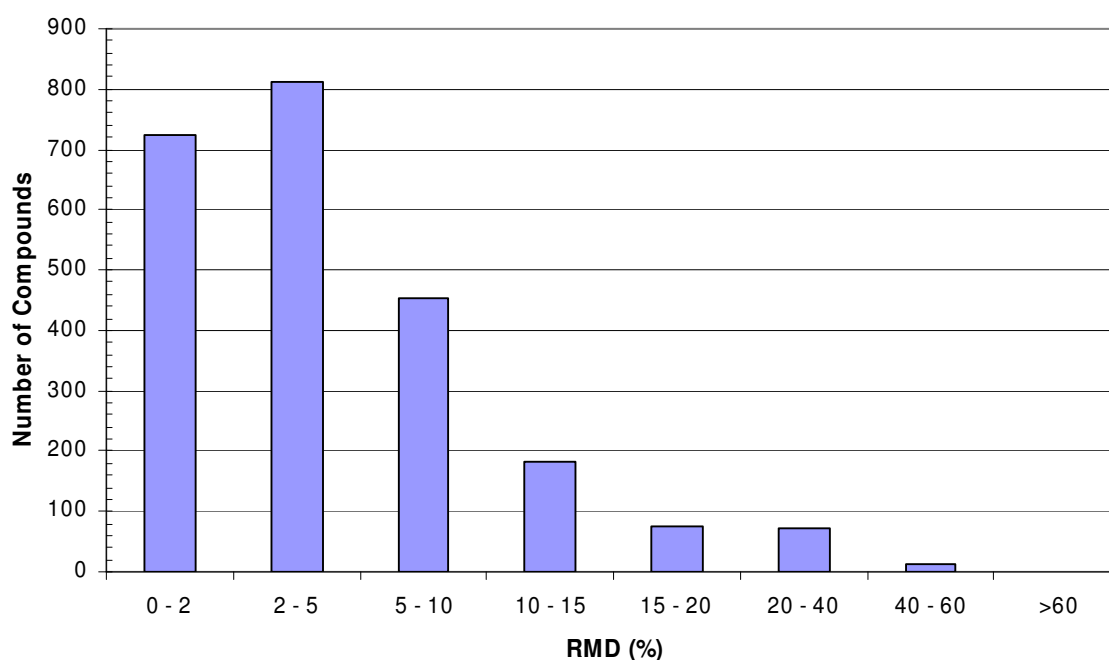


Figure 5.22 Histogram of the vapour pressure relative mean deviation for the compounds in the training set

A total of 212 groups were used to describe the data (both group interaction and group contribution groups), however some groups have only been regressed against 1 or 2 compounds and therefore should only be used as a guide when applied. The group contribution and interaction tables are given in Appendix A. Since there is such a large improvement in the low pressure predictions the model could be used to extrapolate to the normal boiling temperature from low pressure data.

This would be quite a logical application as there are a number of low pressure data available for compounds which do not have a measured boiling point. The drawback of a method such as this is that the extrapolation would only be as good as the data it was extrapolated from, and since low pressure data often has a large amount of scatter this could prove problematic. It is therefore recommended that any boiling point extrapolations be used in conjunction with some of the more conventional boiling point estimation methods available (Rarey et al.⁵ and Nannoolal et al.⁶).

6 CONCLUSIONS

An improved group contribution method has been developed for the prediction of vapour pressures. The group contribution scheme employed by Nannoolal et al. was modified to account for the size dependence of some groups that were found in the development of the model. This allows for the improved prediction of these groups (most noticeable were the aliphatic alcohols, aliphatic carboxylic acids and ketones). The C-parameter used by Nannoolal et al.⁸ was also modified in order to account for a wider range of molecule sizes and to increase the prediction of the vapour pressure model at low pressures. A large improvement in the predictions of the aliphatic alcohol and aliphatic carboxylic acid groups was made by adding in a logarithmic correction term which provides a more physically realistic shape of the curve.

A training set of 2332 compounds (113 888 data points) was used in the development of the model and an average percentage error of 5.0% was found. This compares favourably with the method of Nannoolal et al.⁸ was for 2207 compounds and 111 757 data points had an average percentage error of 6.6 %. The largest improvement was at low pressures (<10 kPa) where the error is almost half of that obtained by the method of Nannoolal et al.⁸. This large improvement is a sensible measure of the performance of the model since the low pressure data are very far from the datum point (which is the normal boiling point).

In order to test the performance of the model (to make sure it was not simply well trained to the data used in the model development) a test set of data, heat of vaporization data at 298 K and solid vapour pressure data were used. The average percentage error for each vapour pressure point in the test set was found to be 7.1 %, which was a bit inflated due to the large amount of scatter that was observed in the test data. The average percentage error for the heat of vaporization data at 298 K was 3.5% which shows that the model parameters are both accurate and physically realistic. For the solid vapour pressure data the average percentage error for each data point was found to be 21.1%; again this was a bit inflated due to the large amount of scatter present in the low pressure data.

7 RECOMMENDATIONS

This work is a continuation of an ongoing project for the prediction of thermophysical properties using group contribution. The properties that have so far been successfully predicted with much acclaim are (all for non-electrolytic organic compounds):

- The normal boiling point
- The critical temperature, pressure and volume
- Viscosity
- Vapour pressure (which was improved in this work)

There are still a huge amount of properties which could be predicted, and with the continual improvement of the software and group definition (through works such as this and Nannoolal et al.) the methods should continue to improve. Some thermophysical properties which could be of interest are:

- Solvent solubility
- Surface tension
- Thermal conductivity
- Melting temperature

The power of group contribution methods is that the model parameters can be predicted from the molecular structure. In this method, for example, it is preferable to have experimental boiling point data but if there is none available it can be predicted from methods currently available^{5,6}. This will obviously affect the accuracy of the prediction but it does mean that a vapour pressure curve can be generated by only knowing the molecular structure.

8 REFERENCES

- ¹ Daubert, T. E., Strengths and Weaknesses of Predictive Methods for Estimating Thermophysical Properties., *J. Chem. Eng. Data* 41 (1996) 942.
- ² Gmehling, J., Rarey, J., Menke, J., Dortmund Data Bank, Oldenburg (2007) <http://www.ddbst.com>.
- ³ Environmental news network, www.enn.com, last accessed 20 July 2007.
- ⁴ Bowen, H. J., Palmer, S. R., Fielder, H. M., Coleman, G., Routledge, P. A., Fone, D. L., Community exposures to chemical incidents: Europe environmental public health surveillance system in development and evaluation of the first., *Jech Online (jech.bmj.com)*, last accessed 20 July 2007.
- ⁵ Rarey, J., Cordes, W., A New Method for the Estimation of the Normal Boiling Point of Non-Electrolyte Organic Compounds., *Fluid Phase Equilibria*, 201 (2002) 409.
- ⁶ Nannoolal, Y., Rarey, J., Ramjugernath, D., Cordes, W., Estimation of Pure Component Properties, Part 1: Estimation of the Normal Boiling Point of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions., *Fluid Phase Equilib.*, 226 (2004) 45-63.
- ⁷ Nannoolal, Y., Rarey, J., Ramjugernath, D., Estimation of Pure Component Properties Part 2: Estimation of Critical Data by Group Contribution., *Fluid Phase Equilib.*, 252 (2007) 1.
- ⁸ Nannoolal, Y., Rarey, J., Ramjugernath, D., Estimation of Pure Component Properties Part 3: Estimation of the Vapor Pressure of Non-Electrolyte Organic Compounds via Group Contributions and Group Interactions., *Fluid Phase Equilib.* submitted for publication.
- ⁹ Wagner, W., A New Correlation Method for Thermodynamic Data Applied to the Vapor Pressure Curve on Argon, Nitrogen and Water. London: IUPAC Thermodynamic Tables Project Centre. (1977).
- ¹⁰ Antoine, C., Vapor Pressure: A New Relation between Vapor Pressure and Temperature., *Compt. Rend.* 107 (1888a) 836.

- ¹¹ Riedel, L., Eine neue universelle Dampfdruckformel Untersuchungen über eine Erweiterung des Theorems der übereinstimmenden Zustände. Teil I., *Chem. Ing. Technol.* 26 (1954) 83.
- ¹² Twu, C. H., Coon, J. E., Cunningham, J. R., A New Generalized Alpha Function for a Cubic Equation of State Part 2. Redlich-Kwong Equation., *Fluid Phase Equilib.* 96 (1994) 61.
- ¹³ Cox, E. R., Hydrocarbon Vapor Pressures., *Ind. Eng. Chem.* 28(5) (1936) 613.
- ¹⁴ Cox, E. R., Pressure-Temperature Chart for Hydrocarbon Vapors', *Ind. Eng. Chem.*, 15(6), (1923) 592.
- ¹⁵ Calingaert, G., Davis, D. G., Pressure-Temperature Charts – Extended Ranges., *Ind. Eng. Chem.* 17 (1925) 1287.
- ¹⁶ DIPPR *Design Institute for Physical Property Data*, New York, AIChE Design Institute for Physical Properties (1992).
- ¹⁷ Myrdal, P. B., Yalkowsky, S. H., Estimating Pure Component Vapor Pressures of Complex Organic Molecules., *Ind. Eng. Chem. Res.* 36 (1997) 2494.
- ¹⁸ Mishra, D. S., Yalkowsky, S. H., Estimation of Vapor Pressure of Some Organic Compounds., *Ind. Eng. Chem. Res.* 30 (1991) 1609.
- ¹⁹ Clegg, S. L., Kleeman, M. J., Griffin, R. J., Seinfeld, J. H., Effects of Uncertainties in the Thermodynamic Properties of Aerosol Components in an Air Quality Model – Part II: Predictions of the Vapour Pressures of Organic Compounds., *Atmos. Chem. Phys. Discuss.* 7(4) (2007) 11049.
- ²⁰ Tu, C., Group Contribution Method for the Estimation of Vapor Pressures., *Fluid Phase Equilib.* 99 (1994) 105.
- ²¹ Lyman, W. J., Reehl, W. F., Rosenblatt, D. H., Handbook of Chemical Property Estimation Methods, McGraw-Hill, Washington DC (1990).
- ²² Fishtine, S. H., Reliable Latent Heats of Vaporization., *Ind. Eng. Chem.* 55 (1963) 47.

- ²³ Voutsas, E., Lampadariou, M., Magoulus, K., Tassios, D., Prediction of Vapor Pressures of Pure Compounds from Knowledge of the Normal Boiling Point Temperature., *Fluid Phase Equilib.* 198 (2002) 81.
- ²⁴ Abrams, D. S., Massaldi, H. A., Prausnitz J. M., Vapor Pressures of Liquids as a Function of Temperature. Two-Parameter Equation Based on Kinetic Theory of Fluids., *Ind. Eng. Chem. Fundam.* 13(3) (1974) 259.
- ²⁵ Moelwyn-Hughes, E. A., The Chemical Statics and Kinetics of Solutions, Academic Press, London (1971).
- ²⁶ Soave, G., Equilibrium Constants from a Modified Redlich-Kwong Equation of State., *Chem. Eng. Sci.* 27 (1972) 1197.
- ²⁷ Coquelet, C., Chapoy, A., Richon, D., Development of a New Alpha Function for the Peng–Robinson Equation of State: Comparative Study of Alpha Function Models for Pure Gases (Natural Gas Components) and Water-Gas Systems., *Int. J. Thermophysics* 25(1) (2004) 133.
- ²⁸ Copeman, T. W., Mathias, P. M., Vapor pressures of pure compounds using the Peng–Robinson equation of state with three different attractive terms., *Fluid Phase Equilib.* 13 (1983) 59.
- ²⁹ Reid, R. C., Prausnitz, J. M., Poling, B. E., The Properties of Gases and Liquids, 4th Ed., McGraw-Hill, USA (1987).
- ³⁰ Eubank, P. T., Wang, X., Saturation Properties from Equations of State., *Ind. Eng. Chem. Res.* 42 (2003) 3838.
- ³¹ Hanif, N. S. M., Shyu, G., Hall, K. R., Eubank, P. T., Areas of J. C. Maxwell for Pure-Component Fluid-Phase Equilibria from Equations of State., *Ind. Eng. Chem. Res.* 35 (1996) 2431.
- ³² Pitzer, K. S., Lippmann, D. Z., Curl, R. F., Huggins, C. M., Petersen, D. E., The Volumetric and Thermodynamic Properties of Fluids. II. Compressibility Factor, Vapor Pressure and Entropy of Vaporization., *J. Am. Chem. Soc.* 77 (1955) 3433.
- ³³ Lee, B. I., Kesler, M. G., A Generalized Thermodynamic Correlation Based on Three-Parameter Corresponding States., *AIChE J.* 21 (1975) 137.

- ³⁴ Thomson, G. W., The Antoine Equation for Vapor-Pressure Data., *Chemical Reviews* 38(7) (1946) 1.
- ³⁵ Ambrose, D., J. The correlation and estimation of vapour pressures IV. Observations on Wagner's method of fitting equations to vapour pressures., *Chem. Thermodynamics* 18 (1986) 45.
- ³⁶ Chase, D. J., Two Models of Vapor Pressure along the Saturation Curve *Ind. Eng. Chem. Res.* 26 (1987) 107.
- ³⁷ Waring, W., Form of a Wide-Range Vapor Pressure Equation., *Ind. Eng. Chem.* 46(4) (1954) 762.
- ³⁸ Liang, C., Gallagher, D. A., QSPR Prediction of Vapor Pressure from Solely Theoretically-Derived Descriptors., *J. Chem. Inf. Comp. Sci.* 38(2) (1998) 321.
- ³⁹ Paul, P. K., Prediction of vapour pressure using descriptors derived from molecular dynamics., *Org. Biomol. Chem.* 3 (2005) 1176.
- ⁴⁰ Ambrose D., Vapor-Pressure Equations. National Physical Laboratory NPL Rep. Chem, 114 (1980).
- ⁴¹ Ledanois, J., Colina, C. M., Santos, J. W., Gonzalez-Mendizabel, D., Olivera-Fuentes, C., *Ind. Eng. Chem. Res.* 36 (1997) 2505.
- ⁴² Ben-Naim, A., Solvation Thermodynamics, Plenum Press (1987).
- ⁴³ Lin, S., Chang, J., Wang, S., Goddard, W. A., Sandler, S. I., Prediction of Vapor Pressures and Enthalpies of Vaporization Using a COSMO Solvation Model., *J. Phys. Chem.* 108 (2004) 7429.
- ⁴⁴ Press et al., Numerical Recipes in Fortran the Art of Scientific Computing, 2nd Ed., Cambridge University Press (1986).
- ⁴⁵ Dennis, J. E., Schnabel R. B., Numerical Methods for Unconstrained Optimization of Nonlinear Equations, Prentice-Hall, New York (1983).

⁴⁶ Lampton, M., Damping-Undamping Strategies for the Levenberg-Marquardt Nonlinear. Least-Squares Method., *Computers in Physics* 11(1) (1997) 110.

⁴⁷ Cordes et al., Software Development in Chemistry 7, Proceeding of the Computer in Chemie 7, Paper presented at the D. Ziessow (Ed.), Springer-Verlag, Berlin (1993).

⁴⁸ Barton, A. F., Handbook of Solubility Parameters and Other Cohesion Parameters, 2nd Ed., CRC Press, USA (1991).

APPENDICES

A GROUP CONTRIBUTION AND INTERACTION TABLES

Table A.1 Group contribution and group interaction values and descriptions

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
-	A	The constant term A	9.42208	-	-	-	-
ALIPHATIC CARBON GROUPS							
1	-CH ₃	Methyl group attached to a non-aromatic non-electronegative atom	-0.00227	2,2-Dimethylbutane	135	No	101
4	-CH ₂ -	CH ₂ in a chain	0.07545	n-Butane	141	No	102
5	>CH-	CH in a chain	0.07099	3-Ethylpentane	144	No	103
6	>C<	C in a chain	-0.04707	2,2,4-Trimethylpentane	146	No	104
2	-CH ₃	Methyl group attached to a non-aromatic electronegative atom	0.13491	N-Methylaniline	132	No	105
7	-CH ₂ -	CH ₂ in a chain attached to an electronegative atom	0.11758	Ethylenediamine	136	No	106
8	-CH<	CH in a chain attached to an electronegative atom	0.08955	5-Ethyl-2-nonanol	137	No	107
9	>C<	C in a chain attached to an electronegative atom	-0.08960	tert-Butanol	138	No	108
29	-CH ₃	Methyl group attached to a ring carbon	-0.07834	Methylcyclohexane	122	No	109
10	-CH ₂ -	CH ₂ in a ring	-0.01350	Cyclohexane	143	No	110
11	>CH-	CH in a ring	0.06029	Methylcyclohexane	145	No	111
12	>C<	C in a ring	0.10842	1,1-Dimethylcyclopentane	147	No	112
131	>CH(r) - C(r)<	CH in a ring bonded to a carbon in a different ring	0.01296	cis-Decahydronaphthalene	110	No	113
136	C(k)-C(r)-3C(r)	Ring carbon attached to 3 other ring carbons and a chain carbon	0.00823	1,3-Dimethyl adamantane	108	No	114
137	C-4C(r)	Ring carbon attached to 4 other ring carbons	0.17344	Spiro[4.5]decane	109	No	115
24	-CH ₂ (r)-en	CH ₂ in a ring attached to an electronegative carbon	0.08201	1,4-Dioxane	142	No	116
14	>CH-	CH in a ring attached to an electronegative atom	0.10344	Cyclopentanol	139	No	117
15	>C<	C in a ring attached to an electronegative atom	-0.12395	Perfluorocyclopentane	140	No	118
139	C(r)_3C(r)_en	Ring carbon bonded to 3 other ring carbons and an en atom	-0.05881	1-Nitroadamantane	107	No	119

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
26	CH ₂ =	Double bonded carbon at the end of a chain/ring	-0.00564	1-Nonene	117	Yes	120
20	-CH=C-	Double bonded carbon in a chain with only 1 carbon neighbour	0.00286	2-Heptene	121	Yes	121
27	-C=C-	Double bonded carbon in a chain with 2 carbon neighbours	0.00475	2-Methyl-2-pentene	118	Yes	122
31	>C=C=C<	C=C=C; cumulated double bonds	0.08031	1,2-Butadiene	112	Yes	123
33	>C=C- C=C<	C=C-C=C (chain); conjugated double bonds (chain)	0.03953	trans-1,3-Pentadiene	114	Yes	124
21	>C=C<	Double bond between carbons in a ring	0.00242	1,3-Cyclopentadiene	123	Yes	125
13	-C=C<	Double bonded carbon in a ring with 2 carbon neighbours	0.02664	1-Methylcyclohexene	120	Yes	126
32	>C=C- C=C<	C=C-C=C (ring); conjugated double bonds (ring)	-0.00045	1,3-Cyclopentadiene	113	Yes	127
134	>C(r)=C(k)	Carbon in a ring double bonded to a carbon outside the chain	0.01488	beta-Pinene	106	Yes	128
25	CH#	Carbon triple bonded to another carbon at the end of a chain	-0.03884	1-Octyne	124	Yes	129
22	-CtC-	Triple bond between 2 carbons in a chain	-0.01491	2-Heptyne	125	Yes	130
34	-C#C-C#C-	C#C-C#C; conjugated triple bonds	0.00977**	2,4-Hexadiyne	111	Yes	131

AROMATIC CARBON GROUPS

3	-CH ₃	Methyl group attached to an aromatic atom	-0.01001	1-Methyl naphthalene	133	No	201
16	-CH(a)<	CH in an aromatic ring	0.01653	Benzene	128	No	202
17	>C(a)<	C in an aromatic ring	0.11192	Propylbenzene	131	No	203
19	=C(a)<	Aromatic carbon attached to three aromatic neighbours	-0.01113	Naphthalene	116	No	204
132	C(a)-C(r) <	Aromatic carbon bonded to a carbon in a ring	0.07303	1,2,3,4-Tetrahydronaphthalene	129	No	205
133	C(a)-C(a)	2 Aromatic carbons chain bonded	0.13938	Benzidine	130	No	206
135	C(a)-r-C(a)	Aromatic carbon bonded to an aromatic carbon in a ring	-0.10703	9H-Fluorene	115	No	207
138	C(a)-C=	Aromatic carbon attached to a double bonded carbon	0.26362	Divinylbenzene	127	No	208
18	>C(a)<	C in an aromatic ring attached to an electronegative atom	0.21038	Aniline	119	No	209

FLUORINE GROUPS

35	F-	Fluorine attached to non-aromatic carbon	0.06101	1-Fluoropentane	92	No	301
38	F-C-1Halo	Fluorine attached to a carbon with one other halogen atom	0.10540	Perfluorocyclopentane	71	No	302

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
39	F-C-2Halo	Fluorine attached to a carbon with two other halogen atoms	0.09402	1,1,1,2-Tetrachloro-2,2-difluoroethane [R112a]	70	No	303
36	F-C=	Fluorine attached to double bonded carbon	0.11304	2-Fluoropropane	69	No	304
148	F-C=(1 Halo)	Fluorine attached to a double bonded carbon with one other halogen atom	0.07386	1,1-Difluoroethene	68	No	305
37	F-C(a)	Fluorine attached to aromatic carbon	-0.07045	Fluorobenzene	91	No	306
116	Fe-Si<	Fluorine attached to a silicon atom	0.21389	Silicon tetrafluoride	1	No	307
CHLORINE GROUPS							
40	Cl-	Chlorine attached to non-aromatic carbon	0.06508	Chloroethane	76	No	401
43	Cl-C-1Halo	Chlorine attached to a carbon with one other halogen atom	0.03302	2-Bromo-2-chloro-1,1,1-trifluoroethane	67	No	402
44	Cl-C-2Halo	Chlorine attached to a carbon with two other halogen atoms	0.05138	1,1,1-Trichloroethane [R140a]	66	No	403
42	Cl-C=	Chlorine attached to double bonded carbon	0.07694	2-Chloro-1,3-butadiene	65	No	404
149	Cl-C=(1 Halo)	Chlorine attached to a double bonded carbon with one other halogen atom	0.12660	Tetrachloroethylene	64	No	405
41	Cl-C(a)	Chlorine attached to aromatic carbon	-0.19988	Chlorobenzene	77	No	406
117	Cl-Si<	Chlorine attached to a silicon atom	-0.00761	Trichlorosilane	2	No	407
113	COCl-	Acid chloride	0.33205	Trichloroacetyl chloride	19	No	408
BROMINE GROUPS							
45	Br-	Bromine attached to non-aromatic carbon	-0.01712	1,2-Dibromoethane	78	No	501
144	Br-C-1Halo	Bromine attached to a carbon with one other halogen atom	0.07426	2-Bromo-2-chloro-1,1,1-trifluoroethane	75	No	502
145	Br-C-2Halo	Bromine attached to a carbon with two other halogen atoms	-0.02594	Tribromomethane [R20B3]	74	No	503
146	Br-C=	Bromine attached to a double bonded carbon	0.15186	Vinyl bromide	73	No	504
46	Br-C(a)	Bromine attached to aromatic carbon	-0.18810	Bromobenzene	79	No	506
118	Br-Si<	Bromine attached to a silicon atom	-0.13300	Silicon tetrabromide	3	No	507
IODINE GROUPS							
47	I-	Iodine attached to carbon	0.02257	Ethyl iodide	61	No	601
119	I-Si<	Iodine attached to a silicon atom	1.30968**	Triiodomethylsilane	4	No	607

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
OXYGEN GROUPS							
53	-COOH (n=<9)	COOH Group attached to a small molecule (n =< 9)	0.81104	Pentanoic acid	25	Yes	701
155	-COOH (n>9)	COOH Group attached to a large molecule (n > 9)	0.08469	Dodecanoic acid		Yes	702
48	C(a)-COOH	Aromatic COOH	2.38380*	Benzoic acid	24	No	703
49	-OH (n=<4)	OH Group attached to a small molecule (n =< 4)	-0.43267	Ethanol	93	Yes	704
153	-OH (n>4)	OH Group attached to a large molecule (n > 4)	-0.04696	5-Ethyl-2-nonanol		Yes	705
50	C(a)-OH	Aromatic OH	0.73847	2-Naphthol	94	No	706
51	-O-	Ether oxygen	0.15049	Diethyl ether	96	No	707
52	-O-	Aromatic oxygen	0.23511	Furan	95	No	708
54	-COO -	Ester in a chain	0.55698	1,3-Benzenedicarboxylic acid dimethyl ester	26	No	709
55	-COO -	Formic acid ester	0.54599	Formic acid ethyl ester	28	No	710
56	-COO -	Ester in a ring (lactones)	0.53129	gamma-Butyrolactone	27	No	711
57	>C=O	Ketone bonded to aromatic ring	0.14889	Acetophenone	58	No	712
58	>C=O	Ketone	-0.03266	Acetone	59	Yes	713
59	-CHO	Aldehyde in chain	0.37695	Acetaldehyde	57	No	714
60	-CHO	Aldehyde attached to an aromatic ring	0.20025	Benzaldehyde	56	No	715
61	O=C(-O-)2	Carbonate diester	0.52435	Carbonic acid dimethyl ester	15	No	716
62	-CO-O-CO-	Anhydrides	0.79451	Acetic anhydride	10	No	717
63	-CO-O-CO-	Cyclic anhydrides with double or aromatic bond	0.95360	Maleic anhydride	9	No	718
64	>(OC2)<	Epoxide	-0.08988	Ethylene oxide	54	Yes	719
65	-O-O-	Peroxides	1.90315*	Di-tert.butyl peroxide	34	No	720
66	-OCOO-	Carbonates O-C=O & -O	0.55927	Propylene carbonate	36	No	721
67	>C=O	Carbonyl(C=O) with S attached to carbon	-0.28344*	Methyl thioacetate	42	No	722
68	>C=O	Urea	0.41594	1,1,3,3-Tetramethyl urea	7	No	723
69	-OCON<	Carbamate	3.68627*	Methyldimethylcarbamate	6	No	724

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
NITROGEN GROUPS							
70	-CONH<	Amide with no substituents	2.15987*	Acetamide	29	No	801
71	-CONH<	Amide with one substituent attached to the nitrogen	1.56627	N-Methylformamide	11	No	802
72	-CONH<	Amide with two substituents attached to the nitrogen	0.46886	N,N-Dimethylformamide (DMF)	12	No	803
73	OCN-	Isocyanate	0.19547	Isocyanic acid methyl ester	31	Yes	804
141	-OCN(a)	IsoCyanate attached to an aromatic carbon	0.49979	Phenyl isocyanate	30	No	805
74	ONC-	Oxime	2.55684	Methyl ethyl ketoxime	32	No	806
75	NO2-	Nitro group attached to a non-aromatic carbon	0.32934	Nitromethane	21	No	807
76	NO2-	Nitro group attached to an aromatic carbon	0.22620	Nitrobenzene	22	No	808
77	NO2-	Nitrite	-0.28378**	Ethyl nitrite	23	No	809
78	-ON=	Isoxazole O-N=C	0.52969*	5-Methyl-4-nitroisoxazole	51	No	810
79	NO3-	Nitrate	0.83647	Ethyl nitrate	14	No	811
80	NH2-	Primary amine attached to non-aromatic carbon/silicon	-0.05329	Ethylenediamine	98	Yes	812
81	NH2-	Primary amine attached to aromatic carbon	0.33218	Aniline	97	No	813
82	-NH-	Secondary amines (chain) attached to carbons/silicons	0.41210	Dibutylamine	102	No	814
86	=N-	Secondary amines (chain) attached to one carbons/silicons via double bond	0.82998	N-Benzylidenemethyl amine	105	No	815
93	-NH-	Secondary amines (ring) attached to carbons/silicons	0.58897	Morpholine	101	No	816
94	-NH-	secondary amines attached to aromatic carbons/silicons	0.30756	N-Methylaniline	33	No	817
84	-N<	Tertiary amine attached to carbons/silicons	-0.22437	N,N-dimethylaniline	104	No	818
92	-N<	tertiary amines attached to aromatic carbon	0.07255	N,N-Dimethylaniline	43	No	819
85	>N<	Nitrogen attached to four carbons	-1.17553*	N,N,N',N'-Tetramethylmethylenediamine	35	No	820
140	C2-N-C(r)	Cyclic tertiary amines	-0.28632	N-Methylpiperidine	103	No	820
83	N=N	Azene N=N	0.00256	Azobenzene	50	No	821
130	>N-N<	A hydrazine functional group	0.96678	Hydrazine	49	No	822
143	N-N_C	Hydrazine with 1 carbon neighbours	0.71254	Phenylhydrazine	48	No	823

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
142	N-N_C2	Hydrazine with 2 carbon neighbours	-0.13598	1,1-Dimethylhydrazine	47	No	824
87	=N-	Aromatic nitrogen in a five-membered ring	0.73669	Oxazole	100	No	825
88	=N-	Aromatic nitrogen in a six-membered ring	0.24836	2-Methylpyridine	99	No	826
89	-CtN (n=<12)	CN Group attached to a small molecule (n=<12)	-0.04339	Acetonitrile	60	Yes	827
154	-C#N (n>12)	CN Group attached to a large molecule (n>12)	0.08072	Tetradecanenitrile		Yes	828
SULFUR GROUPS							
98	-S-S-	Disulfide	0.03557	Dimethyl disulfide	55	No	902
99	-SH	Thiol or mercaptane attached to carbon	-0.02822	1-Propanethiol	80	No	903
100	-S-	Thioether	-0.03261	Ethyl methyl sulfide	81	No	904
101	-S-	Aromatic thioether	-0.11625	Thiazole	82	No	905
102	-SO2-	Sulfone O=S=O	-0.33737*	2,4-Dimethylsulfolane	18	No	906
104	-SO2N<	Sulfon amides, attached to N and to S with 2 double bond O	0.19623	N,N-Dimethyl-methanesulfonamide	38	No	908
105	>S=O	Sulfoxide	0.56077*	Dimethyl sulfoxide	40	No	909
106	SCN-	Isothiocyanat	0.06394	Allyl isothiocyanate	20	No	910
107	>SO3	Sulfate with one oxygen replaced by another atom	0.44583*	Benzenesulfonic acid,ethyl ester	39	No	911
PHOSPHOROUS GROUPS							
95	P(O)O3-	Phosphate triester	0.66806	Methyl diphenyl phosphate	8	No	1001
96	>P<	Phosphine	-0.04275	Triphenylphosphine	46	No	1002
97	PO3-	Phosphite attached to only 3 oxygens,PO3	-0.13750*	Triethoxyphosphine	45	No	1003
METAL GROUPS							
108	>Se<	Selenium	0.48339**	Diselenide, diphenyl	52	No	1101
109	AsCl2-	Arsenic dichloride attached to a carbon	0.36903	Methylarsenic dichloride	17	No	1102
110	>Sn<	Stannane with four carbon neighbours	0.07688	Tetramethylstannane	62	No	1103
111	B(O-)3	Boric acid triester	0.47847	Boric acid trimethyl ester	16	No	1104

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
GERMANIUM GROUPS							
114	GeCl3-	GeCl3 attached to carbon	0.27066	Trichlorosilyl(trichlorogerm yl)methane	13	No	1201
115	>Ge<	Germane with four carbon neighbors	0.34203	Tetramethylgermane	63	No	1202
SILICON GROUPS							
120	-SiH3	Silane group	-0.24084	Butylsilane	84	No	1301
121	-SiH2-	Primary silicon group	0.27568	Trisilane	88	No	1302
122	-SiH<	Secondary silicon group	-0.29038	Triethylsilane	89	No	1303
123	>Si<	Tertiary silicon group	0.03614	Tetraethylsilane	90	No	1304
124	-SiH3	Silane group attached to an electromagnetic atom	0.20445	Monochlorosilane	83	No	1305
127	SiH2	SiH2 attached to electronegative atoms	0.21195	Dichlorosilane	85	No	1306
128	SiH	SiH attached to electronegative atoms	-0.01409	Trichlorosilane	86	No	1307
125	>Si<	Silicon atom bonded to electromagnetic atoms	0.03880	Tetrachlorosilane	87	No	1308
129	CH3-Si	Methyl group attached to a Silicon atom	-0.00451	Tetramethylsilane	134	No	1309
SPECIAL GROUPS							
150	no H	No hydrogen atoms	-0.19373	Perfluoropentane		No	1401
151	one H	One hydrogen atom	-0.04327	2-Bromo-2-chloro-1,1,1- trifluoroethane		No	1402
SIZE DEPENDANT GROUP CONSTANTS							
170		Alkenes group constant	-0.02835			No	2001
171		Alkynes group constant	0.60141			No	2002
172		Ketone group constant	0.71345			No	2003
173		Epoxy group constant	0.91912			No	2004
174		Isocyanate group constant	-0.41147			No	2005
175		Short OH group constant (n=<4)	6.69345			No	2006
176		Long OH group constant (n>4)	5.21138			No	2007
177		Primary amine group constant	0.91589			No	2008

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
179		Short CN group constant (n<12)	0.61103			No	2010
180		Long CN group constant (n>12)	-1.05206			No	2011
181		Short COOH group constant (n<9)	-2.54299			No	2012
182		Long COOH group constant (n>9)	3.92217			No	2013
GROUP INTERACTIONS							
200		Alcohol - Alcohol Interaction	-0.00306			Yes	3001
201		Alcohol - 1 Amine Interaction	-0.26016			No	3002
202		Alcohol - 2 Amine Interaction	-0.50269			No	3003
203		Alcohol - Thiol Interaction	-1.13734*			No	3004
205		Alcohol - Ether Interaction	-0.55743			No	3006
207		Alcohol - Ester Interaction	-1.38632			No	3008
208		Alcohol - Ketone Interaction	-1.38783**			No	3009
210		Alcohol - Cyan Interaction	0.13770*			No	3011
220		1 Amine - 1 Amine Interaction	0.82184			No	3021
221		1 Amine - 2 Amine Interaction	-0.29523**			No	3022
224		1 Amine - Ether Interaction	0.07904			No	3025
226		1 Amine - Ester Interaction	-0.80854*			No	3027
230		1 Amine - Aromatic O Interaction	-0.75793*			No	3031
232		1 Amine - Alcohol (a) Interaction	2.06766*			No	3033
237		1 Amine - Nitro(a) Interaction	0.06456			No	3038
239		2 Amine - 2 Amine Interaction	-0.23506**			No	3040
242		2 Amine - Ether Interaction	-0.05391*			No	3043
257		Thiol - Thiol Interaction	0.46453			No	3058
267		Thiol - Alcohol (a) Interaction	0.24272*			No	3068
274		Carboxy - Carboxy Interaction	-2.42076			No	3075
286		Carboxy - Aromatic S Interaction	-1.67665*			No	3087

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
290		Ether - Ether Interaction	-0.00531			No	3091
291		Ether - Epox Interaction	0.77292			No	3092
292		Ether - Ester Interaction	-0.02349			No	3093
293		Ether - Ketone Interaction	-0.79909			No	3094
294		Ether - ThioEther Interaction	-0.50401*			No	3095
295		Ether - Cyan Interaction	0.01913			No	3096
298		Ether - Alcohol (a) Interaction	-0.60245			No	3099
299		Ether - Aldehyde Interaction	0.00837			No	3100
303		Ether - Nitro(a) Interaction	-0.11088**			No	3104
304		Ether - Iso Cyan(a) Interaction	-1.35234*			No	3105
305		Epox - Epox Interaction	0.13164*			No	3106
319		Ester - Ester Interaction	-0.01683			No	3120
320		Ester - Ketone Interaction	-0.05930			No	3121
322		Ester - Cyan Interaction	0.18719			No	3123
323		Ester - Aromatic O Interaction	0.48770*			No	3124
324		Ester - 6 N Ring Interaction	1.74386**			No	3125
325		Ester - Alcohol (a) Interaction	-1.17782*			No	3126
326		Ester - Aldehyde Interaction	0.12169*			No	3127
332		Ketone - Ketone Interaction	-0.14617			No	3133
334		Ketone - Cyan Interaction	0.67059*			No	3135
344		ThioEther - ThioEther Interaction	-0.19956*			No	3145
355		Cyan - Cyan Interaction	0.26466			No	3156
357		Cyan - 6 N Ring Interaction	0.41394**			No	3158
368		Aromatic O - Aldehyde Interaction	0.22632*			No	3169
371		Aromatic O - 5 N Ring Interaction	-0.04067*			No	3172
374		6 N Ring - 6 N Ring Interaction	0.17913			No	3175

Ink No	Name	Description	dB _i	Example	Prty	Size Dep	Ref No
375		6 N Ring - Alcohol (a) Interaction	-0.81562*			No	3176
382		Alcohol (a) - Alcohol (a) Interaction	-0.04719			No	3183
389		Aldehyde - Aldehyde Interaction	0.56723*			No	3190
395		Iso Cyan - Iso Cyan Interaction	-4.23062*			No	3196
401		Aromatic S - 5 N Ring Interaction	-0.13058**			No	3202
404		5 N Ring - 5 N Ring Interaction	-0.80379*			No	3205
407		Nitro(a) - Nitro(a) Interaction	-0.27452			No	3208
408		Nitro(a) - Iso Cyan(a) Interaction	14.86700			No	3209

Ink No. – The number which is used by the fragmentation program to identify the group

Ref No. – Reference number, used to order the groups since ink no's are very mixed

Size Dep – Groups with yes need to be multiplied by the number of atoms (not incl. hydrogen) in the molecule

Size Dep Group constants – Always have a frequency of 1

Prty – Group Priority – The order in which groups are fragmented – a lower priority is fragmented first

* - group only fitted to data for one compound

** - group only fitted to data for 2 compounds

Table A.2 Group contribution values for the logarithmic correction term

Name	Description	Value
Aliphatic Alcohols		
D	Constant term with a frequency of 1	-4.798
dE _i	The aliphatic alcohol group	6.578
dE _i	The aliphatic alcohol group interaction term	2.888
Aliphatic Carboxylic Acids		
D	Constant term with a frequency of 1	-7.162
dE _i	The aliphatic carboxylic acid group	41.83

B SAMPLE CALCULATIONS

Table B.1 and Table B.2 show sample calculations for two different components. The two examples cover the usage of size dependant groups, group interactions and the logarithmic correction term.

Table B.1 Calculation for the vapour pressure of 1-hexen-3-ol at 389K

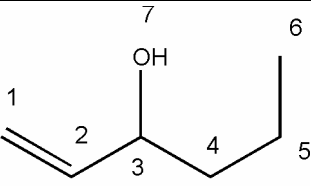
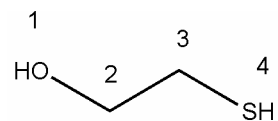
<div style="text-align: center;">  </div> <div style="text-align: center;">1-Hexen-3-ol</div>						T = 389.0 K
						Na = 7
Ink No	atoms	Size Dep	Frequency	Contribution	Total	
1	6	No	1	-0.00227	-0.00227	
4	4,5	No	2	0.07545	0.1509	
8	3	No	1	0.08955	0.08955	
20	2	Yes	1	0.00286	0.02002	
26	1	Yes	1	-0.00564	-0.03948	
153	7	Yes	1	-0.04696	-0.32872	
170	1,2	No	-	-0.02835	-0.02835	
176	7	No	-	5.21138	5.21138	
Total Sum (dB)					5.07303	
D' Parameter						
Nme	atoms	Size Dep	Frequency	Contribution	Total	
D	7	No	-	-4.798	-4.798	
dE	7	No	1	6.578	0.9397	
Total Sum (D')					-3.85829	
<div style="display: flex; justify-content: space-between;"> <div> <p>B' = A + dB</p> <p>B' = 9.42208 + 5.07303</p> <p>B' = 14.4951</p> </div> <div> <p>T_b = 408.2 K</p> <p>C(T_b) = 53.173 K</p> <p>P^s_{exp} = 53.33 kPa</p> </div> </div>						
$P^s = \exp \left(B' \frac{T - T_b}{T - C(T_b)} + D' \ln \left(\frac{T}{T_b} \right) \right) \times 101.325 \text{ kPa}$						
$P^s = \exp \left(14.4951 \frac{389 - 408.2}{389 - 53.173} - 3.8583 \ln \left(\frac{389}{408.2} \right) \right) \times 101.325 \text{ kPa}$						
$P^s = 53.28 \text{ kPa}$						

Table B.2 Calculation for the vapour pressure of 2-mercapto ethanol at 364.8 K



2-Mercapto ethanol

T = 364.8 K

Na = 4

Ink No	atoms	Size Dep	Frequency	Contribution	Total
7	2,3	No	2	0.11758	0.23516
49	1	Yes	1	-0.43267	-1.73068
99	4	No	1	-0.02822	-0.02822
175	1	No	-	6.69345	6.69345
Total Sum (dB)					5.16971
Group Interactions					
203	1,4	No	1	-1.13734	-1.13734
Total Sum (GI)					-1.13734
D' Parameter					
Nme	atoms	Size Dep	Frequency	Contribution	Total
D	1	No	-	-4.798	-4.798
dE	1	No	1	6.578	1.6445
Total Sum (D')					-3.15350

$$B' = A + dB + GI$$

$$B' = 9.42208 + 5.16871 - 1.13734$$

$$B' = 13.45345$$

$$T_b = 422.97 \text{ K}$$

$$C(T_b) = 56.200 \text{ K}$$

$$P_{\text{exp}}^s = 13.27 \text{ kPa}$$

$$P^s = \exp \left(B' \frac{T - T_b}{T - C(T_b)} + D' \ln \left(\frac{T}{T_b} \right) \right) \times 101.325 \text{ kPa}$$

$$P^s = \exp \left(13.4535 \frac{364.6 - 422.97}{364.6 - 56.2} - 3.1535 \ln \left(\frac{364.6}{422.97} \right) \right) \times 101.325 \text{ kPa}$$

$$P^s = 12.68 \text{ kPa}$$

C RIEDEL CALCULATION EXAMPLE

The following example is for the calculation of the Riedel model parameters for benzene (as outlined by Reid et al.²⁹). Benzene has the following properties (obtained from the DDB²):

$$T_c = 562.1 K \quad P_c = 4894 kPa \quad T_b = 353.3 K$$

Riedel defined the parameters in the model (Eq. (2-21)) in terms of the parameter α_c (which is α (Eq. (2-22)) at the critical point) as follows:

$$A = -35Q \quad B = 36Q \quad C = 42Q + \alpha_c \quad D = -Q \quad (C-1)$$

The variable Q was found to have the following dependence on α_c :

$$Q = 0.0838(3.758 - \alpha_c) \quad (C-2)$$

Since α is quite a complex differential the simplest method to calculate α_c is to substitute Eq. (C-1), (C-2) and the normal boiling point (i.e. $P = 1 atm$, $T = T_b$) in to Eq. (2-21) and solve the resulting expression for α_c , which results in the following 2 expressions:

$$\alpha_c = \frac{0.315\psi_b + \ln P_c}{0.0838\psi_b - \ln T_{r_b}} \quad (C-3)$$

$$\psi_b = -35 + \frac{36}{T_{r_b}} + 42 \ln T_{r_b} - T_{r_b}^6 \quad (C-4)$$

where $T_{r_b} = T_b / T_c$. So for benzene $T_{r_b} = 0.629$, then by substituting this into the above equations the following values for the parameters are found:

$$A = 8.939 \quad B = -9.1944 \quad C = -3.9208 \quad D = 0.2554$$

D EQUATIONS FOR $\Delta H/R\Delta Z$

Table D.1 Forms of the equations for $\Delta H/R\Delta Z$ for the various equations used in this dissertation

Equation name	Equation for: $-\frac{\Delta H}{R\Delta Z} = \frac{d \ln P}{d(1/T)}$	Num
Antoine	$\frac{d \ln P}{d(1/T)} = \frac{B}{\left(1 - \frac{C}{T}\right)^2}$	(D-1)
Cox	$\frac{d \ln P}{d(1/T)} = \ln 10 \times \lambda \left(\ln 10 \times E \left(\frac{-2}{T_c^2} T^3 + (1+F) \frac{T^2}{T_c} \right) \left(1 - \frac{T_b}{T} \right) - T \right)$ $\lambda = \exp(\ln 10 \times \log A_c + \ln 10 \times E(1 - T_r)(F - T_r))$	(D-2)
Riedel	$\frac{d \ln P}{d(1/T)} = BT_c - CT - \frac{6D}{T_c^6} T^7$	(D-3)
Myrdal & Yalkowsky	$\frac{d \ln P}{d(1/T)} = B - CT$	(D-4)
Tu	$\frac{d \ln P}{d(1/T)} = B + CT + DT^2$	(D-5)
Watson	$\frac{d \ln P}{d(1/T)} = A[\sigma - \eta]$ $\sigma = 2mT \left(3 - \frac{2T}{T_b} \right)^{m-1} \left((m-1) \left(3 - \frac{3T}{T_b} \right)^{-1} \frac{2T}{T_b} \ln \frac{T_b}{T} + 1 \right)$ $\eta = T_b \left(3 - \frac{2T}{T_b} \right)^m \left(1 + 2m \frac{T}{T_b} \left(3 - 2 \frac{T}{T_b} \right)^{-1} \right)$	(D-6)
Abrams et al.	$\frac{d \ln P}{d(1/T)} = B - CT - BT^2 - 2ET^3$	(D-7)
Lee-Kesler	$\frac{d \ln P}{d(1/T)} = -(6.09648 + \omega 15.6875) T_c + (1.28862 + \omega 13.4721) T$ $- \frac{6}{T_c^6} (0.169347 + \omega 0.43577) T^7$	(D-8)
Wagner	$\frac{d \ln P}{d(1/T)} = \frac{T^2}{T_c(1-\tau)^2} (a + 1.5b\tau^{0.5} - 0.5b\tau^{1.5}$ $+ 3c\tau^2 - 2c\tau^3 - 6d\tau^5 - 5d\tau^6)$	(D-9)

E SOFTWARE SCREENSHOTS

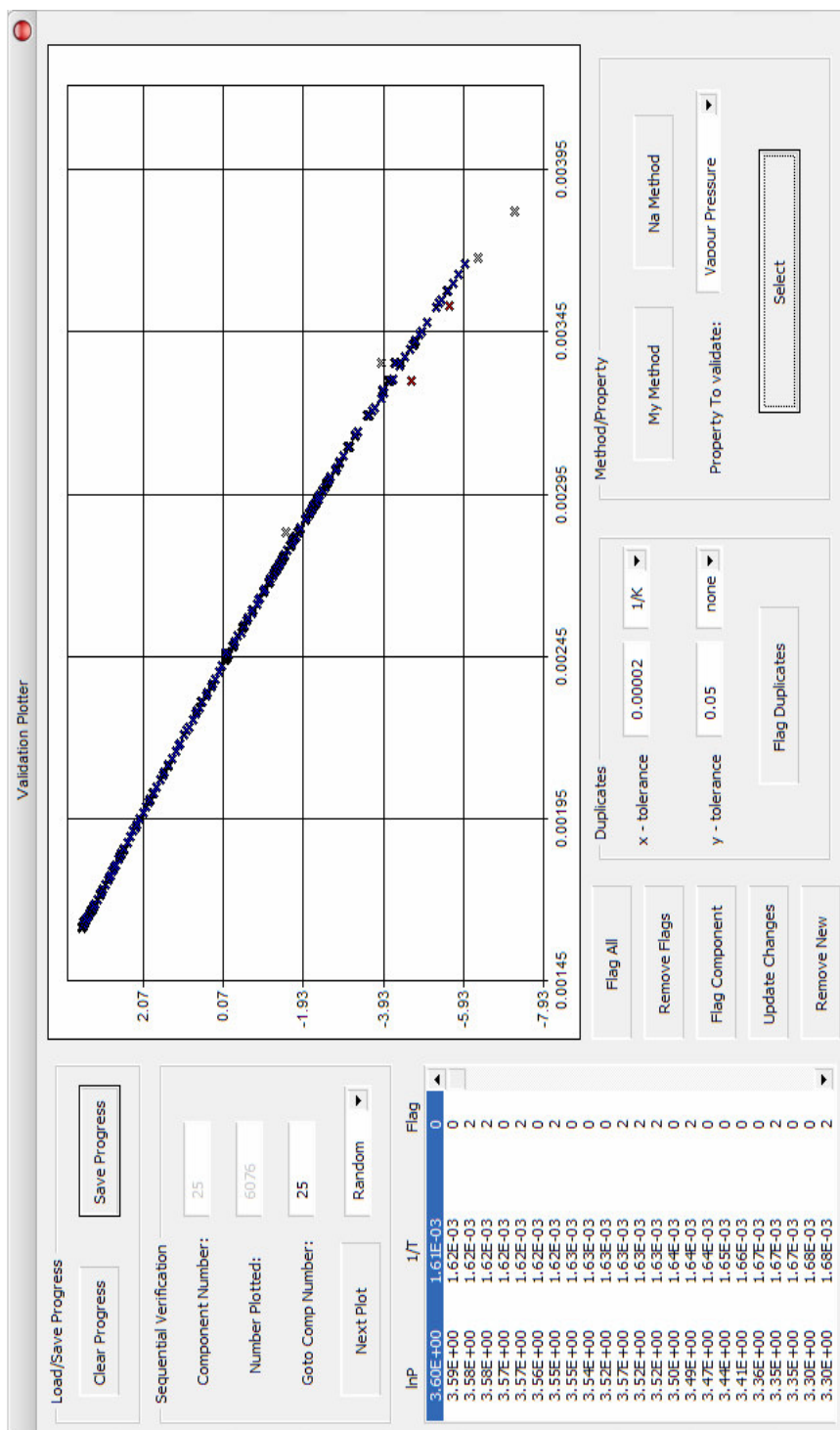


Figure E.1 Screenshot of the GUI used to validate the data – this enabled the fast removal of any obvious outliers

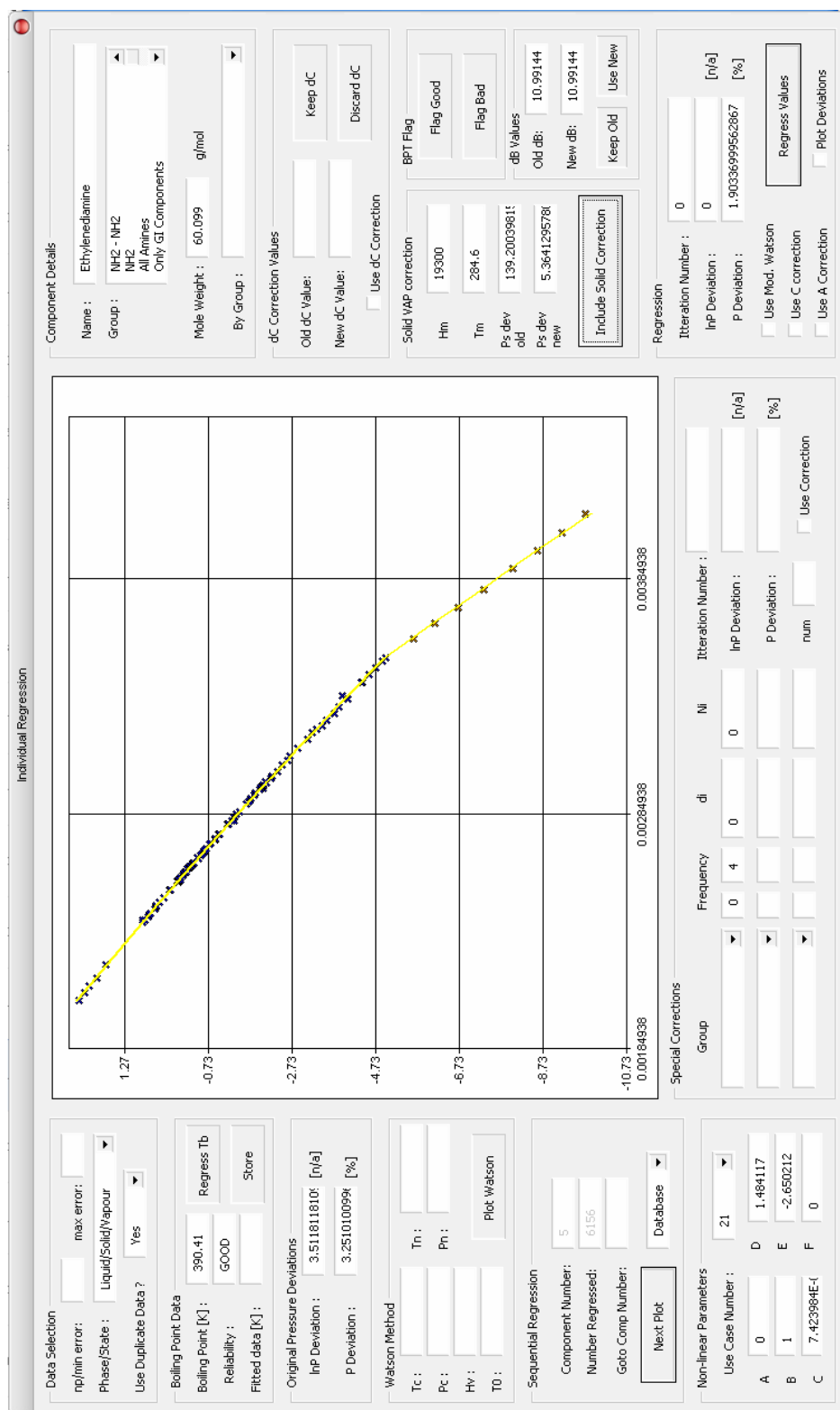


Figure E.2 Screenshot of the GUI used to test the different models – enabled the rapid testing of the forms of the equation that were tested for this work

F CALCULATION OF ΔH_{VAP} FROM EQUATIONS OF STATE

As stated in paragraph 2.2.1 the condition for equilibrium between 2 phases is that the chemical potentials in both phases are equal (Eq. (2-3)). Therefore if we assume a fixed value of temperature ($dT = 0$) and combine Eq. (2-3) and Eq. (2-5) for the transition from a vapour to a liquid (this is possible because for a pure substance $G = \mu$) the following results:

$$\int_{liq}^{vap} V dP = 0 \quad (F-1)$$

Then by using some simple differentiating and rearranging the following expression can be found:

$$V dP = d(PV) - P dV \quad (F-2)$$

Combining Eq. (F-1) and Eq. (F-2):

$$P^s (V^g - V^l) = \int_{V^l}^{V^g} P dV \quad (F-3)$$

Which is equivalent to:

$$\int_{V^l}^{V^g} (P - P^s) dV = 0 \quad (F-4)$$

This is the mathematical expression for the so called Maxwell Equal Area Rule (MEAR). The MEAR is illustrated graphically by Figure F.1 (the van der Waals EOS was purely used for illustrative purposes), and states that for vapour-liquid equilibrium to occur the absolute value of area A_1 must equal the absolute are of area A_2 (this can be easily shown by splitting Eq. (F-4) into 2 separate integrals – by using the additivity of integral intervals). This fact can therefore be used to work out the vapour pressure of a pure substance. Eubank and Wang³⁰ differentiated Eq. (F-3) using Leibnitz's rule (which is simply an application of the fundamental theorem of calculus):

$$P^s \left(\frac{dV_v^s}{dT} - \frac{dV_l^s}{dT} \right) + (V_v^s - V_l^s) \left(\frac{dP^s}{dT} \right) = P^s \left(\frac{dV_v^s}{dT} \right) - P^s \left(\frac{dV_l^s}{dT} \right) + \int_{V_l^s}^{V_v^s} \left(\frac{\partial P^{EOS}}{\partial T} \right) (dV)_T \quad (F-5)$$

Which simplifies to:

$$(V_v^s - V_l^s) \left(\frac{dP^s}{dT} \right) = \int_{V_l^s}^{V_v^s} \left(\frac{\partial P^{EOS}}{\partial T} \right)_V (dV)_T \quad (F-6)$$

The expression on the left hand side of Eq. (F-6) can be replaced by Eq. (2-11) resulting in the following expression of the heat of vaporization:

$$\Delta H_{vap} = T \int_{V_l^s}^{V_v^s} \left(\frac{\partial P^{EOS}}{\partial T} \right)_V (dV)_T \quad (F-7)$$

Therefore with a pressure explicit EOS such as the SRK (Eq. (F-8)) the heat of vaporization can be given by Eq. (F-9).

$$P = \left(\frac{RT}{V-b} \right) - \left(\frac{a}{V(V+b)} \right) \quad (F-8)$$

$$\frac{\Delta H_{vap}}{RT} = \ln \left(\frac{V_v^s - b}{V_l^s - b} \right) + \frac{da}{dT} \frac{1}{bR} \ln \left(\frac{(V_v^s + b)V_l^s}{(V_l^s + b)V_v^s} \right) \quad (F-9)$$

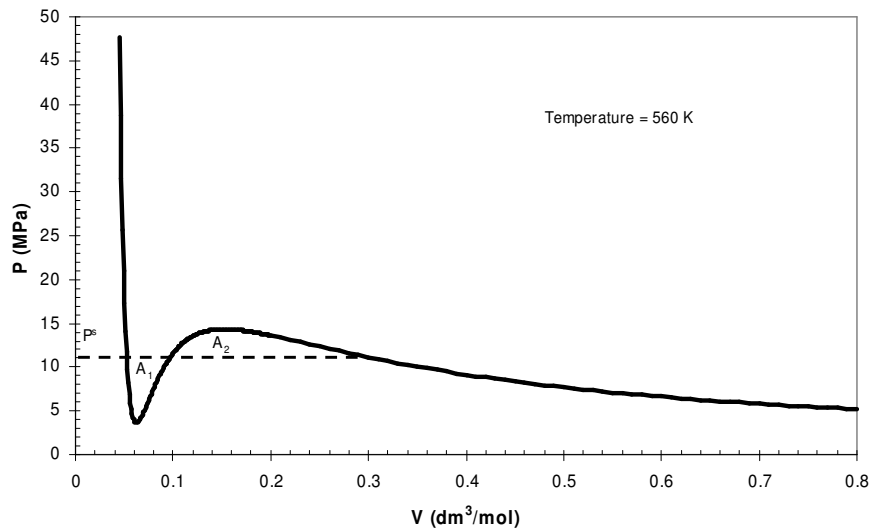


Figure F.1 P vs. V for water at 560 K as given by the van der Waals EOS

Eq. (F-9) is a generalised expression and depending on the alpha function that is used da/dT will change, for the Soave alpha function Eq. (F-9) becomes:

$$\frac{\Delta H_{vap}}{RT} = \ln\left(\frac{V_v^s - b}{V_l^s - b}\right) - \left(\frac{0.42748m\sqrt{\alpha}}{0.08664\sqrt{T_r}}\right) \ln\left(\frac{(V_v^s + b)V_l^s}{(V_l^s + b)V_v^s}\right) \quad (F-10)$$

Eubank and Wang³⁰ suggest using the Racket equation (Eq. (F-11)) for the saturated liquid volume and the virial equation of state truncated to the third term (Eq. (F-12)) for the saturated liquid volume.

$$V_l^s = \frac{RT_c}{P_c} Z_c^{1+(1-T_r)^{2/7}} \quad (F-11)$$

$$\frac{P^s V_v^s}{RT} = 1 + \frac{B}{V_v^s} + \frac{C}{V_v^s} \quad (F-12)$$

However this introduces more variables and makes the equation less widely applicable (while data (and correlations) for the 2nd virial coefficient is widely available in the literature, data (and correlations) for the 3rd virial coefficient are more scarce). Therefore if it is assumed that the SRK EOS provides a good estimate of the vapour pressure (which in the case of benzene it does) we can rearrange Eq. (F-10) in terms of the compressibility factor of the liquid and the vapour and get these values from the SRK EOS.

$$\frac{\Delta H_{vap}}{RT} = \ln\left(\frac{Z_v - B}{Z_l - B}\right) - \left(\frac{0.42748m\sqrt{\alpha}}{0.08664\sqrt{T_r}}\right) \ln\left(\frac{(Z_v + B)Z_l}{(Z_l + B)Z_v}\right) \quad (F-13)$$

As shown in Figure F.2, this equation can provide a moderately good prediction of the heat of vaporization, however when showing the behaviour of $\Delta H_{vap}/(R\Delta Z_{vap})$ (Figure 2.13) it falls away quite noticeably. The reason for this is that even though Figure F.2 seems like a good representation, closer inspection reveals that above 500 K ($T_r \approx 0.9$) the SRK prediction is an underestimate of the data and this is consistent with what is observed in Figure 2.13. A superior representation of the heat of vaporization can be found by using the alpha function of Twu et al.¹². Figure F.3 shows how the Twu alpha function gives a better-quality fit up to $T_r \approx 0.97$ and then it deviates quite significantly.

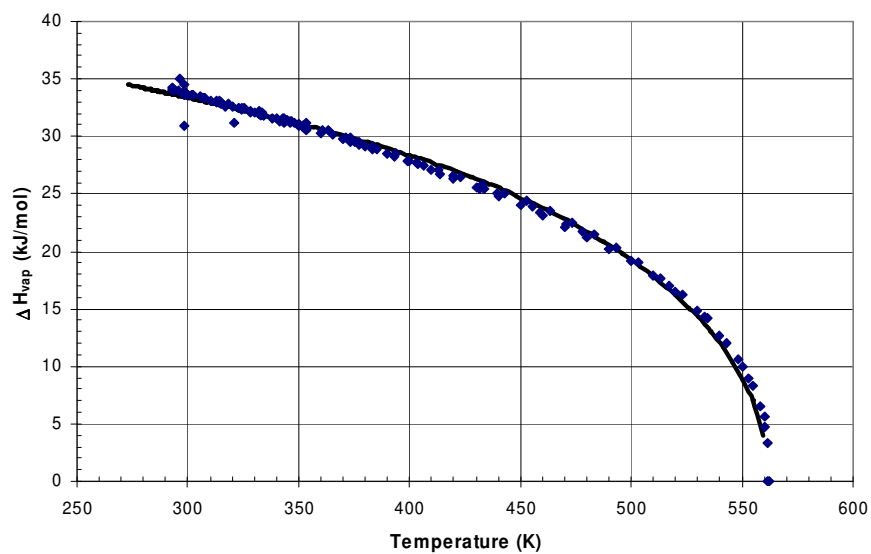


Figure F.2 Heat of vaporization of benzene as a function of temperature (♦ – data from the DDB², — SRK prediction Eq. (F-13))

$$\frac{da}{dT} \frac{1}{bR} = \left(\frac{0.42748}{0.08664} \right) \left(NT_r^{MN-N-1} \exp(1 - T_r^{MN}) ((m-1) - LMT_r^{MN}) \right) \quad (\text{F-14})$$

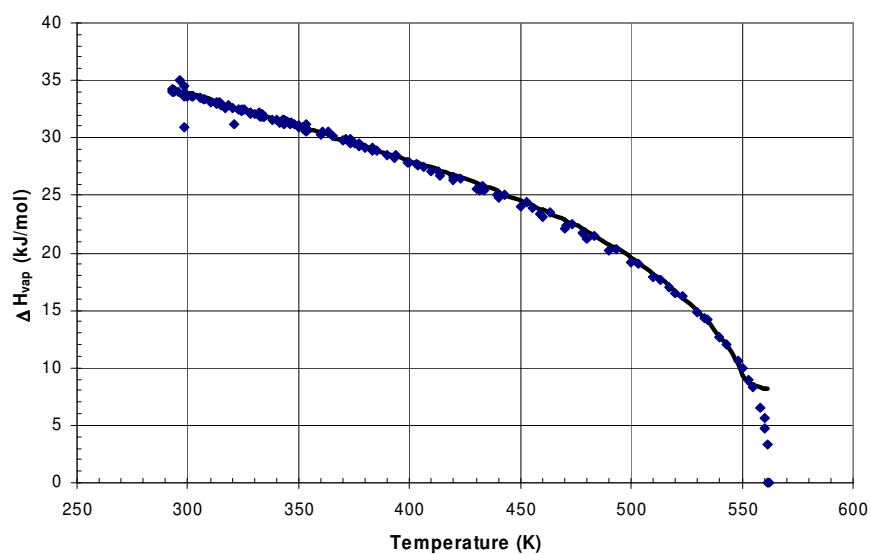
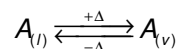


Figure F.3 Heat of vaporization of benzene as a function of temperature (♦ – data from the DDB², — Twu SRK prediction Eq.(F-14) and Eq. (F-9) in the form of Eq. (F-13))

G CHANGE IN THE CHEMICAL POTENTIAL

Consider the vaporization/condensation of substance A (the +/- Δ refer to the fact that vaporization of A is endothermic and condensation of A is exothermic):



The vaporization can be expressed in terms of the change in Gibbs free energy as follows:

$$\Delta G = \mu_A^v - \mu_A^l \quad (G-1)$$

The change in Gibbs free energy can also be written in terms equilibrium constant (K):

$$\Delta G = -RT \ln K \quad (G-2)$$

The temperature dependence of the equilibrium constant is given by Eq. (G-3) and is known as the van't Hoff equation.

$$\frac{d \ln K}{dT} = \frac{\Delta H}{RT^2} \quad (G-3)$$

Therefore for the vaporization of A, $\Delta H = \Delta H_{\text{vap}}$ and since vaporization is endothermic ΔH_{vap} will always be a positive number. This means that an increase of temperature will cause an increase in the equilibrium constant (Eq. (G-3)) this in turn will cause ΔG to be negative (Eq. (G-2)) which therefore results in Eq. (2-2).

H FURTHER NOTES ON DATA VALIDATION AND DATA USED

As mentioned in paragraph 3.2 the data was validated by plotting the data on the inverse temperature log pressure axes and removing any outliers. The main reason why such a method was chosen is that it provides a (relatively) quick method of fairly accurately screening the data. Unfortunately the problem with this method is that there is no real way to distinguish between good and poor (this is not to say that the data is very inaccurate but that it is slightly less accurate than some of the high precision data available in the database) data where there is a slight scatter. [For examples of this type of scatter see Figure 2.16 and Figure 5.1 – while it is clear that the data is good there is still a small scatter which is very difficult to rectify]

Data could be screened on the basis of the journal which it comes from, however this may not always be fair, since experimental errors are not limited to one journal and not another, similarly with the authors. Another possibility is to produce deviation plots; however this rests on the premise that there is some basis from which to take the deviation. No doubt for some of the more common compounds in the database (benzene, hexane etc.) such a basis does exist (e.g. using accurate Wagner parameters to generate a curve) and could be used fairly successfully, however these compounds are in the minority and are then subject to external factors (the accuracy of the parameters). Another approach that was tried was to take the deviation of the data relative to the line joining the highest and the lowest value. This approach made it very difficult to gauge if the proper shape of the curve was being maintained. Also for some compounds the endpoints of the dataset were erroneous which lead to garbage being produced. For this reason the only viable option which presented itself was to use the $1/T$ vs. $\ln P$ method that was used.

It is for this reason that the errors in section 2.2 may seem a little more inflated that they should be. For example one would expect the Wagner equation to be able to reproduce the benzene or hexane curves to within a fraction of a percent. Indeed, which a set of highly accurate data, it can. However since all the errors reported for this work and the work of Nannoolal were taken relative to the data that is contained in the database, these same data were used to provide a fairer, more useful comparison.